

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 05:43:30 ; Search time 1815.27 Seconds  
(without alignments)  
3677.450 Million cell updates/sec

Title: US-09-824-286-5

Perfect score: 319

Sequence: 1 GATATCGTATGACCCAGTC.....AGGACCAAGCTGAGATCT 319

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenDbml:\*

- 1: gb\_da:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match Length	ID	Description
-----			

1	292.4	91.7	324	10	MUSNC41	M83538 Mus musculu
2	291.4	91.3	321	10	AF163743	AF163743 Mus muscu
3	289.8	90.8	393	10	AF004328	AF004328 Mus muscu
4	288.2	90.3	353	10	AY058908	AY058908 Mus muscu
5	283.8	89.0	324	10	MUSICKV50	L16827 Mus musculu
6	283.4	88.8	462	10	MMICKV	X05877 Mus musculu
7	281.8	88.3	909	6	AR007979	AR007979 Sequence
8	281.8	88.3	909	6	AR058994	AR058994 Sequence
9	281.8	88.3	909	6	I23444	I23444 Sequence 1
10	281.8	88.3	909	6	I43364	I43364 Sequence 7
11	281.8	88.3	909	6	I61419	I61419 Sequence 7
12	281.8	88.3	909	6	I56033	I56033 Sequence 7
13	279.6	87.6	318	10	MUSICMRNAJ	L21025 Mus musculu
14	278.6	87.3	318	10	MUSICLAF	M36251 Mouse Ig ka
15	276.4	86.6	321	10	MUSICKGR	M41120 Mouse Ig ac
16	274.8	86.1	324	10	MUSICKV25	L16823 Mus musculu
17	273.8	85.8	321	10	MMNPRCLB6	X70097 M.musculus
18	272.2	85.3	324	10	MUSICKRAD	L20956 Mus musculu
19	272.2	85.3	324	10	MUSICKV77	L16838 Mus musculu
20	272.2	85.3	456	10	MUSICKCNK	M19914 Mouse Ig re
21	271.8	85.2	324	10	MUSICKV08	L16817 Mus musculu
22	271.6	85.1	321	6	I32985	I32985 Sequence 39
23	271.6	85.1	324	10	MUSICKV57	L16832 Mus musculu
24	271.6	85.1	324	10	MUSICKV59	L16834 Mus musculu
25	271.6	85.1	327	10	AF192257	AF192257 Mus muscu
26	271.6	85.1	360	6	I32994	I32994 Sequence 48
27	271.6	85.1	360	23	E09038	E09038 cDNA encodi
28	270.6	84.8	321	10	MMNPRCLB1	X70095 M.musculus
29	270.6	84.8	324	10	AF087027	AF087027 Mus muscu
30	270.6	84.8	381	10	AF001496	AF001496 Mus muscu
31	270.2	84.7	324	10	MUSICKV11	L16815 Mus musculu
32	269.6	84.5	324	10	MUSICKV31	L16825 Mus musculu
33	269.6	84.5	721	10	MMU235962	AJ235962 Mus muscu
34	269.2	84.4	324	10	MUSICKV17B	L16821 Mus musculu
35	269	84.3	384	10	MUSICKA	L41875 Mus musculu
36	268.8	84.3	313	10	MUSICKAQ	M64165 Mus musculu
37	267.8	83.9	320	10	MUSICKV6A	L19553 Mus musculu
38	267.4	83.8	381	10	AF045512	AF045512 Mus muscu
39	266.8	83.6	303	10	MUSIGLAF	M36250 Mouse Ig ka
40	266.8	83.6	321	10	MMALCVAR23	X00898 M.musculus
41	266.8	83.6	362	10	AF276281	AF276281 Mus muscu
42	266.4	83.5	324	10	MUSICKV56	L16830 Mus musculu
43	266	83.4	324	10	MUSICKV63	L16836 Mus musculu
44	265.8	83.3	321	10	MUSVHMAB	L32698 Mouse reacr
45	265.8	83.3	341	10	MUSICKCKP	M17161 Mouse Ig ka

## ALIGNMENTS

RESULT 1	MUSNC41	MUSNC41	324 bp	mRNA	linear	ROD 27-APR-1993
LOCUS	MUSNC41	Mus musculus NC41 mRNA, partial cds.				
DEFINITION	M83538					
ACCESSION	M83538.1	GI:200017				
VERSION						
KEYWORDS		antibody light chain variable domain VL.				
SOURCE		Mus musculus Spleen cDNA to mRNA.				
ORGANISM		Mus musculus				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS		Hudson,P.J., Tulip,W., Varghese,J., Laver,W.G., Webster,R. and Colman,P.M.				
TITLE		Refined structure of the influenza virus N9 Neuraminidase-NC41 Fab complex				
JOURNAL		Unpublished (1991)				
FEATURES		Location/Qualifiers				
SOURCE		1..324				
		/organism="Mus musculus"				
		/db_xref="taxon:10090"				
		/tissue_type="Spleen"				
		1..324				
gene						

```

/ gene="NC41"
<1. .>324

```

KLEIKR"			
BASE COUNT	87 a	87 c	
ORIGIN			
		77 g	73 t

Query Match	91.7%;	Score 292.4;	DB 10;	Length 324;
Best Local Similarity	-95.0%;	Pred. No. 3.1e-88;		
Matches 302; Conservative	0;	Mismatches 16;	Indels 0;	Gaps 0;

**OY**    1 GATTTCTGTAAGTCCCGACGTCTGCACAATTCAATGGTCACATCAGTAGGAGACGATATCAC C 60  
         ||| | | | | | | | | | | | | | | | | | |  
**Dd**     1 GCATTGTGTCAGCCCCGACTCTCACAAATTCATGTGCCAATCATCGTGGAAGACAGGGTGACC 60  
         ||||| | | | | | | | | | | | | | | | |

Accession	Sequence	Length
QY	ATCACCCTGCAGGCCAGTCCAGCATGTCTACTCTGTAGCCTGGTATCAACAAAAACCA	120
61	ATCACCCTGCAGGCCAGTCCAGCATGTCTACTCTGTAGCCTGGTATCAACAAAAACCA	120
Db	ATCACCCTGCAGGCCAGTCCAGCATGTCTACTCTGTAGCCTGGTATCAACAAAAACCA	120

Oy 121 GGGCAATCTCCTAAACTTCTGATTTAACGTGGGCATCCACC CGG CACACTGAGATGCCCTCAT 180

Dd 121 GGCAATCTCCTAAACTGATTATTACTGGGCATCCACC CGG CACATGAGATGCCCTCAT 180

0y 181 CGCTTCACAGGCAGTGGATCTGTGGACAGATTATTACTCCACCATCAGCAGTGCGAGGCT 240  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
db 183 CCTTCGCCAGGACTGGATCTGTGGACAGATTATTACTCCACCATCAGCAGTGCGAGGCT 240

[illegible]

QY 301 GGGACCAAGCTGAGATC 318  
II IIIIIIII II  
II IIIIIIII II

## RESULT 2

LOCUS	321 bp	linear	ROD 20-JUL-2001
DEFINITION	Mus musculus mab 24.1.2 immunoglobulin light chain variable region mRNA, partial cds.		
AF163743			

```

house mouse.
.
KEYWORDS
VERSION
AF163743.1 GI:5690292
M163743.1
SOURCE

```

ORGANISM	REFERENCE
Mus musculus	1 (pages 1 to 321)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	

TITLE  
 AUTHORS  
 Merten, N.M., Galvin, J.E., Adderson, E.E. and Cunningham, M.W.  
 Molecular analysis of cross-reactive anti-myoDin/anti-streptococcal  
 mouse monoclonal antibodies  
 J. Biol. Chem. 265: 1125-1132 (1990)

MEDLINE 21179651  
 PUBMED 11282394  
 REFERENCE 2 (Pages 1 to 321)

**TITLE** Direct Submission  
**JOURNAL** Submitted (03-JUN-1999) Microbiology and Immunology, Oklahoma University Health Sciences Center, 940 St. Young Blvd, Oklahoma

```

FEATURES
source      City, OK 73190, USA
            Location/Qualifiers
            1. .321

```

```

/organism="Mus musculus"
/strain="BALB/c"

```

```

/dbc_xref="taxon:10090"
/note="mAb 24.1.2"
<1. .>321

```

	KLELK <sup>a</sup>			
BASE COUNT	86 a	86 c	76 g	73 t
ORIGIN				

Query Match	91.3%;	Score 291.4;	DB 10;	Length 321;
Best Local Similarity	95.0%;	Pred. No. 6.8e-88;		
Matches 301; Conservative	0;	Mismatches 16;	Indels 0;	Gaps 0;

**Dy**    1 GATATCGTAAATGACCCAGTCCTCAACAATTCAATGTGCCCATCAGTAGAGACAGTATACC 60  
       || | | | | | | | | | | | | | | | |  
**Db**    1 GACATTGTGATGACCCCAGTCTCACAAATTCATGTCCCATCATGTAGAGACAGGGTGAC 60

Qy	Qy
61	ATACACCTCCAGAGCCAGTCAGAGTGTGTACTACTGCTGTAGCTGTGTATCAACAAAAACA 120
61	ATACACCTCCAGAGCCAGTCAGAGTGTGTACTACTGCTGTAGCTGTGTATCAACAAAAACA 120
Db	ATACACCTCCAGAGCCAGTCAGAGTGTGTACTACTGCTGTAGCTGTGTATCAACAAAAACA 120

Qy	121	GGGCAATCTCTTAAACTCTGTATTTACTGGGCATCACCACCCGGCACACTGTGAGTCCCTGAT	180
	121	GGGCAATCTCTCTTAAACTCTGTATTTACTGGGCATCACCACCCGGCACACTGTGAGTCCCTGAT	180
Db	121	GGGCAATCTCTCTTAAACTCTGTATTTACTGGGCATCACCACCCGGCACACTGTGAGTCCCTGAT	180

181 CGCTTCACAGGAGTGGATCTGTGGACAGATTATCTTCACCATCAGCAGTGTGACAGGCT 240  
 OY |||||  
 181 CGCTTCACAGGAGTGGATCTGTGGACAGATTATCTTCACCATCAGCAGTGTGACAGGCT 240  
 DB |||||

[illegible]

QY 301 GGGACCAAGCTGGAGAT 317  
|||||

### RESULT 3

LOCUS	393 bp	DNA	linear	ROD 24-MAR-1998
DEFINITION	AF004328			
	Mus musculus Ig kappa light chain variable region, monoclonal 19c2a antibody IIB4 specific to haemagglutinin of Influenza A virus,			

gene, partial cds.  
ACCESSION AF004328  
VERSION AF004328.1 GI:2209218  
KEYWORDS

SOURCE	ORGANISM
house mouse.	Chordata:
Mus musculus	Cranialia: Vertebrata: Euteleostomi:
	Sciuromorphi: Muridae: Murinae: Mus
	murinae: Eutheria: Rodentia

REFERENCE  
AUTHORS  
1 (pages 1 to 393)  
Betakova, T., Vareckova, E., Kostolansky, F., Mucha, V. and  
Daniels, R.S.

epitope of influenza virus haemagglutinin elicits biologically significant immune responses  
J. Gen. Virol. 79 (Pt 3), 461-470 (1998)

REFERENCE  
AUTHORS  
TITLE  
2 (bases 1 to 393)  
Danilets, R.S., Kostoliansky, F., Vareckova, E., Russ, G. and Betakova, T.  
Direct Submission

**JOURNAL** Submitted (16-MAY-1997) Department of Molecular Biology and Morphogenesis of Viruses, Institute of Virology, Slovak Academy of Sciences, Dubravska cesta 9, Bratislava 84246, Slovak Republic

```

      REMOVED          LOCATION/QUALITY
SOURCE              1. .393

```

```

CDS      1. .336
         /note="WR17 VL"

```

RESULT	5
MUSIGKV50	
LOCUS	324 bp mRNA linear ROD 29-OCT-1994
DEFINITION	Mus musculus Ig rearranged kappa-chain gene V19-JI region.
ACCESSION	L16827
KEYWORDS	L16827.1 GI:293626
SOURCE	V-region; complementarily determining region; immunoglobulin light chain; kappa-immunoglobulin; processed gene.
ORGANISM	Mus musculus (strain NZB/W F1) mRNA. Mus musculus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 324) Lustgarten,D.L., Kavalier,J., Gerhard,W. and Scharif,M.D. The response to a foreign antigen in the autoimmune NZB/W F1 murine strain Unpublished (1993)
JOURNAL FEATURES	Location/Qualifiers
source	1..324 /organism="Mus musculus" /strain="NZB/W F1" /db_xref="taxon:10090" /cell_line="BW5-50" /cell_type="hybridoma" 1..285 /gene="Igk-V19" /standard_name="V19" 1..324 /gene="Igk-V19" 70..102 /gene="Igk-V19" /standard_name="CDR1"
misc_feature	

```
misc_feature 148..168
                /gene="Igk-V19"
                /standard_name="CDR2"
misc_feature 265..291
                /gene="Igk-V19"
                /standard_name="CDR3"
J-segment 286..324
                /gene="Igk-V19"
                /standard_name="J1"
BASE COUNT 82 a 81 c 79 g 74 t 8 others
ORIGIN
Query Match 89.0%; Score 283.8; DB 10; Length 324;
Best Local Similarity 92.5%; Pred. No. 2.6e-85;
Matches 294; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAAATTCATGTCACATGATGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GACATTTGATGATGACCCAGTCTCACAAATTCATGTCACATGATGAGACAGTATCAC 60
OY 61 ATCACCCTCAAGGCCAGTCAGAGTGTGACTACTGCTGAGCTGATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATCACCCTCAAGGCCAGTCAGAGTGTGACTACTGCTGAGCTGATCAACAAAACCA 120
OY 121 GGGCAATCTCTTAACCTTGATTTACTGTGAGCATCCACCGGCACACTGGAGTCCCTGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCAATCTCTTAACCTTGATTTACTGTGAGCATCCACCGGCACACTGGAGTCCCTGAT 180
OY 181 CGCTTCACAGGAGTGATGATCTGGACAGATTTACTCTCACCATAGAGAGTGGAGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 CGCTTCACAGGAGTGATGATCTGGACAGATTTACTCTCACCATAGAGAGTGGAGGCT 240
OY 241 GAAGACCTGGCACTTTATTTACTGTGACAGCAATTTATTCCTCGTGACGTTGGTGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GAAGACCTGGCACTTTATTTACTGTGACAGCAATTTATTCCTCGTGACGTTGGTGA 300
OY 301 GGGACCAAGCTGGAGATC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GGGACCAAGCTGGAGATC 318

RESULT 6
LOCUS MMIGKV 462 bp mRNA linear ROD 04-AUG-1992
DEFINITION Mouse mRNA for immunoglobulin kappa variable region.
ACCESSION X05877.1 GI:52195
VERSION X05877.1 GI:52195
KEYWORDS constant region; Ig kappa light chain; variable region.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 462)
AUTHORS Grant,F.J.
TITLE Direct Submision
JOURNAL Submitted (31-Aug-1987) Dr. Francis Grant, Zymogenetics, Inc., 2121
North 35th Street, Seattle, WA 98103, USA
REFERENCE 2 (bases 1 to 462)
AUTHORS Grant,F.J., Levin,S.D., Gilbert,T. and Kindsvogel,W.
TITLE Improved RNA sequencing method to determine immunoglobulin mRNA
sequence
JOURNAL Nucleic Acids Res. 15 (13), 5496 (1987)
MEDLINE 87260030
FEATURES
    source
        Location/Qualifiers
            1..462
                /organism="Mus musculus"
                /strain="ML-05"
                /db_xref="taxon:10090"
                /cell_type="hybridoma"
                /note="variable region (AA 1-137)"
CDS
    7..>462
```

```
/codon_start=1
/product="V(kappa) gene product"
/protein_id="CAA29301.1"
/db_xref="GI:52196"
/translation="MHRTSMGKIMKESQIAQFVFLMISGVDSDIVMOSHREKSTSV
GDRVSTICKASQDVTYAVSHYQOKPGOSPRLIFIMASRIHRHGVDFRTGSSGTDYTL
TISSVQADLDLALYTCQDHYSTPLTFGAGTKLELRADAPVYSIFPSSSE"
misc_feature 418..>462
                /note="constant region (AA 138-152)"
BASE COUNT 114 a 121 c 112 g 115 t
ORIGIN
Query Match 88.8%; Score 283.4; DB 10; Length 462;
Best Local Similarity 93.4%; Pred. No. 3.6e-85;
Matches 296; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAAATTCATGTCACATGATGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GACATTTGATGATGACCCAGTCTCACAAATTCATGTCACATGATGAGACAGTATCAC 153
OY 61 ATCACCCTCAAGGCCAGTCAGAGTGTGACTACTGCTGAGCTGATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 ATCACCCTCAAGGCCAGTCAGAGTGTGACTACTGCTGAGCTGATCAACAAAACCA 213
OY 121 GGGCAATCTCTTAACCTTGATTTACTGTGAGCATCCACCGGCACACTGGAGTCCCTGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GGGCAATCTCTTAACCTTGATTTACTGTGAGCATCCACCGGCACACTGGAGTCCCTGAT 180
OY 214 GGGCAATCTCTTAACCTTGATTTACTGTGAGCATCCACCGGCACACTGGAGTCCCTGAT 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 GGGCAATCTCTTAACCTTGATTTACTGTGAGCATCCACCGGCACACTGGAGTCCCTGAT 273
OY 181 CGCTTCACAGGAGTGATGATCTGGACAGATTTACTCTCACCATAGAGAGTGGAGGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 CGCTTCACAGGAGTGATGATCTGGACAGATTTACTCTCACCATAGAGAGTGGAGGCT 333
OY 241 GAAGACCTGGCACTTTATTTACTGTGACAGCAATTTATTCCTCGTGACGTTGGTGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 GAAGACCTGGCACTTTATTTACTGTGACAGCAATTTATTCCTCGTGACGTTGGTGA 393
OY 301 GGGACCAAGCTGGAGAT 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 GGGACCAAGCTGGAGAT 410

RESULT 7
LOCUS AR007979 909 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5753204.
ACCESSION AR007979
VERSION AR007979.1 GI:3967088
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE 1 (bases 1 to 909)
AUTHORS Huston,J.S., Houston,L.L., Ring,D.B. and Oppermann,H.
TITLE Biosynthetic binding proteins for immunotargeting
JOURNAL Patent: US 5753204-A 1 19-MAY-1998;
FEATURES
    source
        Location/Qualifiers
            1..909
                /organism="unknown"
BASE COUNT 229 a 218 c 246 g 216 t
ORIGIN
Query Match 88.3%; Score 281.8; DB 6; Length 909;
Best Local Similarity 93.1%; Pred. No. 1.3e-84;
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
OY 1 GATATCGTAATGACCCAGTCTCACAAATTCATGTCACATGATGAGACAGTATCAC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 GATATCGTAATGACCCAGTCTCTTAATTCATGTCACATGATGAGACAGTATCAC 467
OY 61 ATCACCCTCAAGGCCAGTCAGAGTGTGACTACTGCTGAGCTGATCAACAAAACCA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



Db 468 ATCTCTGCAAGGCCAGTGCAGGATGTGATCTGCTAGCTGTATCAACAAAACCA 527  
Qy 121 GGGCAATCTCTAAACTCTGATTTTACAGGATCCACCGGACACAGGAGTCCCTGAT 180  
Db 528 GGGCAATCTCTAAACTCTGATTTTACAGGATCCACCGGACACAGGAGTCCCTGAT 587  
Qy 181 CGCTTCACAGGACAGTGTGAGACAGATTTATCTCTACCATCAGCAGTGTGAGGCT 240  
Db 588 CGCTTCACAGGACAGTGTGAGACAGATTTATCTCTACCATCAGCAGTGTGAGGCT 647  
Qy 241 GAAGACCTGGCAGCTTTTATCTGTCCAGCAACATTAATACCTCCGTGAGCTTGGTGA 300  
Db 648 GAAGACCTGGCAGCTTTTATCTGTCCAGCAACATTAATAGTCCGTACACGTTCCGAGGG 707  
Qy 301 GGGACCAAGCTGGAGAT 317  
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 8  
AR058994  
LOCUS AR058994 909 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 1 from patent US 5837846.  
ACCESSION AR058994  
VERSION AR058994.1 GI:5984571  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS Huston,J.S., Houston,L.L., Ring,D.B. and Oppermann,H.  
TITLE Biosynthetic binding proteins for immuno-targeting  
JOURNAL Patent: US 5837846-A 1 17-NOV-1998;  
FEATURES  
source 1..909  
location/Qualifiers  
BASE COUNT 229 a 218 c 246 g 216 t  
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;  
Best Local Similarity 93.1%; Pred. No. 1.3e-84;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 GATATCGTAATGACCCAGTCTCAAAATTCATGTCACATGAGTAGAGATACACC 60  
Db 408 GATATCGTAATGACCCAGTCTCTAAATTCATGTCACATGAGTAGAGAGGTCAGC 467  
Qy 61 ATCACTGCAAGGCCAGTGCAGATGTGACTGCTGTAGCCTGGTATCAACAAAACCA 120  
Db 468 ATCTCTGCAAGGCCAGTGCAGATGTGACTGCTGTAGCCTGGTATCAACAAAACCA 527  
Qy 121 GGGCAATCTCTAAACTCTGATTTTACAGGATCCACCGGACACAGGAGTCCCTGAT 180  
Db 528 GGGCAATCTCTAAACTCTGATTTTACAGGATCCACCGGACACAGGAGTCCCTGAT 587  
Qy 181 CGCTTCACAGGACAGTGTGAGACAGATTTATCTCTACCATCAGCAGTGTGAGGCT 240  
Db 588 CGCTTCACAGGACAGTGTGAGACAGATTTATCTCTACCATCAGCAGTGTGAGGCT 647  
Qy 241 GAAGACCTGGCAGCTTTTATCTGTCCAGCAACATTAATACCTCCGTGAGCTTGGTGA 300  
Db 648 GAAGACCTGGCAGCTTTTATCTGTCCAGCAACATTAATAGTCCGTACACGTTCCGAGGG 707  
Qy 301 GGGACCAAGCTGGAGAT 317  
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 9  
LOCUS 123444 909 bp DNA linear PAT 07-OCT-1996  
DEFINITION Sequence 1 from patent US 5534254.

ACCESSION 123444  
VERSION 123444.1 GI:1603314  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS Huston,J.S., Houston,L.L., Ring,D.B. and Oppermann,H.  
TITLE Biosynthetic binding proteins for immuno-targeting  
JOURNAL Patent: US 5534254-A 1 09-OCT-1996;  
FEATURES  
source 1..909  
location/Qualifiers  
BASE COUNT 229 a 218 c 246 g 216 t  
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;  
Best Local Similarity 93.1%; Pred. No. 1.3e-84;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 GATATCGTAATGACCCAGTCTCAAAATTCATGTCACATGAGTAGAGAGATACACC 60  
Db 408 GATATCGTAATGACCCAGTCTCTAAATTCATGTCACATGAGTAGAGAGGTCAGC 467  
Qy 61 ATCACTGCAAGGCCAGTGCAGATGTGACTGCTGTAGCCTGGTATCAACAAAACCA 120  
Db 468 ATCTCTGCAAGGCCAGTGCAGATGTGACTGCTGTAGCCTGGTATCAACAAAACCA 527  
Qy 121 GGGCAATCTCTAAACTCTGATTTTACAGGATCCACCGGACACAGTGTGAGGCT 180  
Db 528 GGGCAATCTCTAAACTCTGATTTTACAGGATCCACCGGACACAGTGTGAGGCT 587  
Qy 181 CGCTTCACAGGACAGTGTGAGACAGATTTATCTCTACCATCAGCAGTGTGAGGCT 240  
Db 588 CGCTTCACAGGACAGTGTGAGACAGATTTATCTCTACCATCAGCAGTGTGAGGCT 647  
Qy 241 GAAGACCTGGCAGCTTTTATCTGTCCAGCAACATTAATACCTCCGTGAGCTTGGTGA 300  
Db 648 GAAGACCTGGCAGCTTTTATCTGTCCAGCAACATTAATAGTCCGTACACGTTCCGAGGG 707  
Qy 301 GGGACCAAGCTGGAGAT 317  
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 10  
LOCUS 143364 909 bp DNA linear PAT 07-OCT-1997  
DEFINITION Sequence 7 from patent US 5631158.  
ACCESSION 143364  
VERSION 143364.1 GI:2468608  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS Dorel,H. and Oppermann,H.  
TITLE Methods and compositions for high protein production from non-native DNA  
JOURNAL Patent: US 5631158-A 7 20-MAY-1997;  
FEATURES  
source 1..909  
location/Qualifiers  
BASE COUNT 229 a 218 c 246 g 216 t  
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;  
Best Local Similarity 93.1%; Pred. No. 1.3e-84;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 GATATCGTAATGACCCAGTCTCAAAATTCATGTCACATGAGTAGAGAGATACACC 60

||||| 408 GATATCGTATGACCCAGTCTCTTAATTCATGTCACGTGAGGAGACGGGTGACG 467  
QY 61 ATCAGCTCAAGGCGCAGTGTGACTGCTGTAGCTGTATCAACAAAACCA 120  
Db 468 ATCTCTGCAAGGCGCAGTGTGACTGCTGTAGCTGTATCAACAAAACCA 527  
QY 121 GGGCAATCTCTTAATCTCTGATTACTGGGATCCACCGGACAGTGTGAGTCCCTGAT 180  
Db 528 GGGCAATCTCTTAATCTCTGATTACTGGGATCCACCGGACAGTGTGAGTCCCTGAT 587  
QY 181 CGCTTCACAGGAGTGTGATCTGGGACATTTACTCTCACCATAGCAGTGTGAGGCT 240  
Db 588 CGCTTCACAGGAGTGTGATCTGGGACATTTACTCTCACCATAGCAGTGTGAGGCT 647  
QY 241 GAAGACCTGGCATTATTACTGTGAGCAACATTTATCACTCCGTGAGCTGTGAGGA 300  
Db 648 GAAGACCTGGCATTATTACTGTGAGCAACATTTATAGAGTCCGTACACGTTCCGAGGG 707  
QY 301 GGGACCAAGCTGGAGAT 317  
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 11  
LOCUS 161419  
DEFINITION Sequence 7 from patent US 5658763.  
ACCESSION 161419  
VERSION 161419.1 GI:2479367  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS Doral, H. and Oppermann, H.  
TITLE Methods and compositions for high protein production from non-native DNA  
JOURNAL Patent: US 5658763-A 7 19-AUG-1997;  
FEATURES Location/Qualifiers  
source 1..909  
BASE COUNT 229 a 218 c 246 g 216 t  
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;  
Best Local Similarity 93.1%; Pred. No. 1.3e-84;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 GATATCGTATGACCCAGTGTGACTGCTGTAGCTGTATCAACAAAACCA 120  
Db 408 GATATCGTATGACCCAGTGTGACTGCTGTAGCTGTATCAACAAAACCA 467  
QY 61 ATCAGCTCAAGGCGCAGTGTGACTGCTGTATCAACAAAACCA 120  
Db 468 ATCTCTGCAAGGCGCAGTGTGACTGCTGTATCAACAAAACCA 527  
QY 121 GGGCAATCTCTTAATCTCTGATTACTGGGATCCACCGGACAGTGTGAGTCCCTGAT 180  
Db 528 GGGCAATCTCTTAATCTCTGATTACTGGGATCCACCGGACAGTGTGAGTCCCTGAT 587  
QY 181 CGCTTCACAGGAGTGTGATCTGGGACATTTACTCTCACCATAGCAGTGTGAGGCT 240  
Db 588 CGCTTCACAGGAGTGTGATCTGGGACATTTACTCTCACCATAGCAGTGTGAGGCT 647  
QY 241 GAAGACCTGGCATTATTACTGTGAGCAACATTTATCACTCCGTGAGCTGTGAGGA 300  
Db 648 GAAGACCTGGCATTATTACTGTGAGCAACATTTATAGAGTCCGTACACGTTCCGAGGG 707  
QY 301 GGGACCAAGCTGGAGAT 317  
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 12  
LOCUS 196033  
DEFINITION Sequence 7 from patent US 5733782.  
ACCESSION 196033  
VERSION 196033.1 GI:3940503  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS Doral, H. and Oppermann, H.  
TITLE Methods and compositions for high protein production from non-native DNA  
JOURNAL Patent: US 5733782-A 7 31-MAR-1998;  
FEATURES Location/Qualifiers  
source 1..909  
BASE COUNT 229 a 218 c 246 g 216 t  
ORIGIN

Query Match 88.3%; Score 281.8; DB 6; Length 909;  
Best Local Similarity 93.1%; Pred. No. 1.3e-84;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 1 GATATCGTATGACCCAGTGTGACTGCTGTAGCTGTATCAACAAAACCA 120  
Db 408 GATATCGTATGACCCAGTGTGACTGCTGTATCAACAAAACCA 467  
QY 61 ATCAGCTCAAGGCGCAGTGTGACTGCTGTATCAACAAAACCA 120  
Db 468 ATCTCTGCAAGGCGCAGTGTGACTGCTGTATCAACAAAACCA 527  
QY 121 GGGCAATCTCTTAATCTCTGATTACTGGGATCCACCGGACAGTGTGAGTCCCTGAT 180  
Db 528 GGGCAATCTCTTAATCTCTGATTACTGGGATCCACCGGACAGTGTGAGTCCCTGAT 587  
QY 181 CGCTTCACAGGAGTGTGATCTGGGACATTTACTCTCACCATAGCAGTGTGAGGCT 240  
Db 588 CGCTTCACAGGAGTGTGATCTGGGACATTTACTCTCACCATAGCAGTGTGAGGCT 647  
QY 241 GAAGACCTGGCATTATTACTGTGAGCAACATTTATCACTCCGTGAGCTGTGAGGA 300  
Db 648 GAAGACCTGGCATTATTACTGTGAGCAACATTTATAGAGTCCGTACACGTTCCGAGGG 707  
QY 301 GGGACCAAGCTGGAGAT 317  
Db 708 GGGACCAAGCTGGAGAT 724

RESULT 13  
LOCUS MUS16MRNAJ  
DEFINITION Mus musculus immunoglobulin light chain (IgL) mRNA, VJ region, partial cds.  
ACCESSION L21025  
VERSION L21025.1 GI:309391  
KEYWORDS Ig light chain; V-region; V-region; immunoglobulin.  
SOURCE Mus musculus (strain BALB/c, sub-species domesticus) male adult hyperimmunized spleen mRNA.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 318)  
AUTHORS Iwasaki, Y., Takabatake, H. and Ferrone, S.  
TITLE Anti-idiotypic monoclonal antibody variable region gene  
JOURNAL Unpublished (1993)  
FEATURES Location/Qualifiers  
source 1..318  
/organism="Mus musculus"





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 06:41:25 ; Search time 363.1 Seconds  
(without alignments)  
1508.388 Million cell updates/sec

Title: US-09-824-286-5  
Perfect score: 319  
Sequence: 1 GATATCGTATGACCCAGTC.....AGGACCACTGAGATCT 319

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	315.8	99.0	319	AAT97440	Monoclonal antibody
2	281.8	88.3	909	AAT36878	741F8 anti-c-erbB-
3	281.8	88.3	909	AAT91837	Anti-c-erbB-2 sing
4	281.8	88.3	909	AAT78879	Single chain antib
5	281.8	88.3	909	AAT21796	741F8 anti-c-erbB-
6	281.8	88.3	909	AAV22598	DNA encoding singl
7	281.8	88.3	909	AAV63397	Anti-c-erbB-2 sFv
8	273.8	85.8	2001	AA046088	Sequence encoding
9	271.6	85.1	360	AA090434	DNA encoding anti-

10	268.4	84.1	321	AA090424	DNA encoding Immun
11	265.8	83.3	394	AAV39345	Mouse anti-HM1.24
12	265.8	83.3	394	AAV36286	CDNA encoding L ch
13	261.2	81.9	321	AA090423	DNA encoding Immun
14	259.8	81.4	303	ABA94232	Mouse 04 antibody
15	259.6	81.4	420	AA090433	DNA encoding anti-
16	257.2	80.6	705	AAV97063	DNA encoding TRA-8
17	256.2	80.3	380	AAA29700	CC92 VL region nuc
18	256.2	80.3	380	AAZ40721	Antibody CC92 high
19	256.2	80.3	380	AA050350	Mouse DNA encoding
20	256.2	80.3	381	AA04265	Colon Cancer monoc
21	254	79.6	411	AAT73901	Variable light sub
22	254	79.6	411	AAV28409	Antibody 24-31 hum
23	253	79.3	882	AAV40022	CDNA insert of PK1
24	251.4	78.8	321	AA065552	Mouse anti-HIV m3
25	251.4	78.8	321	AA070370	Anti HIV antibody
26	251.4	78.8	882	AAV40023	mRNA encoding Kapp
27	250.8	78.6	411	AA005557	Sequence encoding
28	250.6	78.6	321	AAV25487	Human mAb 12B1 lig
29	249.8	78.3	321	AAV294887	Tie2 receptor anta
30	248.6	77.9	748	AAT05781	Erbb-2-specific sc
31	248.2	77.8	321	AAT58265	Lead binding Mab 2
32	248.2	77.8	321	AAV81046	Murine 708 VL DNA
33	248.2	77.8	324	AAV80813	DNA encoding light
34	248.2	77.8	324	AA004592	Sequence encoding
35	248.2	77.8	387	AA027132	Murine antibody variab
36	248.2	77.8	387	AA099034	Light chain variab
37	248.2	77.8	407	AA028254	Murine antibody Fv
38	247.6	77.6	382	AA094484	VL FRP5. Mus musc
39	247.6	77.6	382	AAV38599	Plasmid pUC-M21-V
40	247.6	77.6	794	AAT42034	Chimeric human/mu
41	247.6	77.6	1371	AAZ50584	Plasmid pMW152-5 f
42	247.6	77.6	1389	AAZ50585	M79scFv-interleuk1
43	247.6	77.6	1479	AAT05783	M79scFv-interleuk1
44	247.6	77.6	1692	AAT29409	Leader scFv(FRP5):
45	247.6	77.6	1862	AAT29413	scFv(FRP5)-DETA-DG

## ALIGNMENTS

RESULT 1	
AAT97440	AA097440 standard; cDNA; 319 BP.
XX	XX
AC	AAT97440;
XX	XX
DT	21-MAY-1998 (first entry)
XX	XX
DE	Monoclonal antibody CP.B8 light chain variable region cDNA.
XX	XX
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; CP.B8; Immunological disease;;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;;
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;
SS.	SS.
XX	XX
OS	Mus musculus.
XX	XX
PN	MO9743416-A1.
XX	XX
PD	20-NOV-1997.
XX	XX
PF	09-MAY-1997; 97WO-US07870.
XX	XX
PR	10-MAY-1996; 96US-0017466.
XX	XX
PA	(BIOJ ) BIOGEN INC.
XX	XX
PI	Benjamin CD, Burkly LC, Hession C, Whitty A;

DR WPI: 1998-008885/01.  
DR P-PSDB: AAW31647.  
PT Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
XX  
PS Claim 21: Page 81-82; 111pp: English.  
XX  
XX This cDNA sequence codes for the light chain variable region (see  
CC AAW31647) of monoclonal antibody (Mab) C9.B8, which is produced  
CC by a hybridoma deposited as ATCC 12107, and which is specific for  
CC the gamma constant (gc) chain (see AAW31646) of human cytokine  
CC receptors. A Mab having complementarity determining regions  
CC encoded by the VL or VH (see AAT07441) cDNA is claimed. The  
CC invention provides compositions and methods for inhibiting cytokine  
CC signaling using gc chain blocking agents for the treatment of  
CC immunological diseases such as myasthenia gravis, rheumatoid  
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
CC allergy, asthma, parasitic infection, graft vs. host disease or  
CC psoriasis. Nucleic acids encoding the gc-blocking agents can be  
CC used in gene therapy for the same purposes. A preferred gc  
CC blocking agent in Mab C9.B8 or its Fab fragment.  
XX  
XX Sequence 319 BP; 86 A; 87 C; 71 G; 75 T; 0 other;

Query Match 99.0%; Score 315.8; DB 19; Length 319;  
Best Local Similarity 99.4%; Pred. No. 1.4e-93;  
Matches 317; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGATACACC 60  
Db 1 gatacgtaatgacccagctcacaatcctcacaatcagtagagacagatcacc 60  
OY 61 ATCACTGCAAGGCGAGTGGATGCTACTGCTAGCCTGGATCAACAAAACCA 120  
Db 61 atcacctgcaagcgagtcgagtagtgcactgctgtagctgtaglatacaaaaacca 120  
OY 121 GGGCAATCTCCTAACTTGTATTACTGGGCATCCACCGGACACTGGAGTCCCTGAT 180  
Db 121 gggcaatctcctaaacttctgtattactggtgagtcacccggcacactggtccctgat 180  
OY 181 CGCTTCACAGGAGTGGATCTGGACAGATTACTCTCACCATACAGAGTGCAGGCT 240  
Db 181 cgcttcacagagtcgagtcgagtagtgcacttaccatcagcagtgtagcagct 240  
OY 241 GAAGACCTGGCACTTTATTACTGTCAGCAACAATTATATCACTCCGTGGACGTTGCGTGA 300  
Db 241 gaagacctggcactttattactgtcagcaacaattatatactccgtggagcgttcggtgga 300  
OY 301 GGGACCAAGCTGGAGATCT 319  
Db 301 gggaccaagctggagatct 319

RESULT 2  
AAT36878 standard; cDNA: 909 BP.  
XX  
XX AAT36878:  
XX  
XX 29-OCT-1996 (first entry)  
XX  
XX 741F8 anti-c-erbB-2 two single chain Fv construct.  
XX  
XX 741F8: anti-c-erbB-2; monoclonal antibody; single chain Fv; sfv;  
XX construct; polypeptide linker; C-terminal amino acid sequence;  
XX in vivo imaging; drug targeting experiment; homodimer;  
XX increased; binding avidity; tissue retention time; ss.  
XX  
XX Homo sapiens.  
OS

XX  
FH Key Location/Qualifiers  
FT CDS 3..755  
FT CDS /\*tag= a  
FT  
XX  
XX US5534254-A.  
XX  
XX 09-JUL-1996.  
XX  
XX 06-FEB-1992; 92US-0831967.  
XX  
XX 07-OCT-1993; 93US-0133804.  
XX  
XX 06-FEB-1992; 92US-0831967.  
XX  
XX (CHIR ) CHIRON CORP.  
XX (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
XX Houston LI, Huston JS, Oppermann H, Ring DB;  
XX  
XX WPI: 1996-333194/33.  
XX  
XX P-PSDB: AAW02278.  
XX  
XX Compns. confg. antigen-targeting antibody fragment constructs -  
XX comprising dimer of single-chain Fv fragments  
XX  
XX Example 1: Columns 27-28; 30pp: English.

The variable heavy (VH) and variable light (VL) genes of the 741F8  
CC anti-c-erbB-2 monoclonal antibody (Mab), were isolated from the  
CC cDNA of the parental 741F8 hybridoma line. A two single chain Fv  
CC (sfv) gene was constructed by connecting the VH and VL genes with a  
CC DNA sequence encoding a polypeptide linker. A synthetic DNA duplex  
CC encoding the C-terminal amino acid sequence, (gly)4-Cys was  
CC inserted, and the resulting 741F8 anti-c-erbB-2 two sfv inserted  
CC into an expression vector. The resulting gene, the present  
CC sequence, was transformed into E. coli, and protein expression  
CC induced by the addn. of IPTG to the culture medium.  
CC A compsn. comprising a carrier and the 2 sfv protein prod. can be  
CC used for in vivo imaging, and drug targeting experiments. The  
CC 2 sfv protein prod. is a homodimer, in which both fragments target  
CC the same antigen, therefore giving greater binding avidity and  
CC longer tissue retention times, compared to individual sfv protein  
CC prod. fragments.  
XX  
XX Sequence 909 BP; 229 A; 218 C; 246 G; 216 T; 0 other;

Query Match 88.3%; Score 281.8; DB 17; Length 909;  
Best Local Similarity 93.1%; Pred. No. 3e-82;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 1 GATATCGTAATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGATACACC 60  
Db 408 gatacgtaatgacccagctcacaatcctcacaatcagtagagacaggtcagc 467  
OY 61 ATCACTGCAAGGCGAGTGGATGCTACTGCTAGCCTGGATCAACAAAACCA 120  
Db 61 atcacctgcaagcgagtcgagtagtgcacttaccatcagcagtgtagcagct 467  
OY 121 GGGCAATCTCCTAACTTGTATTACTGGGCATCCACCGGACACTGGAGTCCCTGAT 180  
Db 121 gggcaatctcctaaacttctgtattactggtgagtcacccggcacactggtccctgat 180  
OY 181 CGCTTCACAGGAGTGGATCTGGACAGATTACTCTCACCATACAGAGTGCAGGCT 240  
Db 181 cgcttcacagagtcgagtcgagtagtgcacttaccatcagcagtgtagcagct 647  
OY 241 GAAGACCTGGCACTTTATTACTGTCAGCAACAATTATATCACTCCGTGGACGTTGCGTGA 300  
Db 241 gaagacctggcactttattactgtcagcaacaattatagagtcgtagacggttcggtgga 300  
OY 301 GGGACCAAGCTGGAGATCT 317  
Db 301 gggaccaagctggagatct 317

DB	708	gggaccacagctggagat	724
RESULT	3		
ID	AA791837	standard; CDNA; 909 BP.	
XX	AA791837		
XX	AA791837;		
DT	07-JAN-1998	(first entry)	
XX			
DE	Anti-c-erbB-2 single chain antibody 741F8 scfv gene.		
XX			
KW	Adenovirus; E1A; transactivator; transcription activator; stimulate;		
KW	expression vector; single-chain binding protein; VAI; enhance; PCR;		
KW	translation; production; immortal; eukaryotic cell; scfv; primer;		
KW	single-chain antibody fragment; imaging; tumour; breast cancer;		
KW	ovarian cancer; c-erbB-2 antigen; digoxin intoxication; ss.		
XX			
OS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	2..755	
FT		/tag= a	
XX			
PN	US5658763-A.		
XX			
PD	19-AUG-1997.		
XX			
PF	25-OCT-1993;	93US-0143498.	
XX			
PR	25-OCT-1993;	93US-0143498.	
XX			
PR	05-JUN-1995;	95US-0463675.	
XX			
PA	(CREA-) CREATIVE BIOMOLECULES INC.		
XX			
PI	Dorai H, Oppermann H;		
DR	WPI; 1997-424235/39.		
XX	P-PSDB; AAW29261.		
XX			
PT	Producing single chain binding protein in immortalised eukaryotic		
PT	cells - which comprise protein coding sequences, a transcription		
PT	activator and translation promotion sequences, provides high		
PT	expression at low copy number		
XX			
XX	Example 2; Column 29-32; 24p; English.		
XX			
CC	This sequence encodes a single chain anti-c-erbB-2 antibody (scfv)		
CC	derived from hybridoma 741F8. The protein product includes a C-terminal		
CC	tail Gly4-Cys in some constructs. Single-chain antibodies can be		
CC	produced using a novel method which comprises culturing an immortalised		
CC	eukaryotic cell having transfected DNA sequences (encoding the protein		
CC	of interest), integrated into its genome. In particular expression		
CC	effector vectors containing a non-native reporter DNA (encoding the scfv)		
CC	and viral sequences to promote transcription and translation (e.g. the		
CC	adenovirus E1A and VAI genes as shown in AA791831 and AA791834		
CC	respectively) are used. The scfv that is produced, when properly folded,		
CC	has a structure with mono- or bi-functional binding activity. The method		
CC	is especially used to produce single-chain antibody fragments (scfv),		
CC	e.g. for imaging tumours or delivering therapeutic agents to them,		
CC	particularly breast and ovarian cancers that express the c-erbB-2		
CC	antigen. Other scfv are used in model studies and for treating digoxin		
XX	intoxication.		
XX			
XX	Sequence 909 BP; 229 A; 218 C; 246 G; 216 T; 0 other;		

Query Match	88.3%	Score 281.8;	DB 18;	Length 909;
Best Local Similarity	93.1%;	Pred. No. 3e-82;		
Matches 295;	Conservative 0;	Mismatches 22;	Indels 0;	Gaps 0;
1	GATATCGTAATGACCAATCTCACAAATTATCATGTCCACATCAGTAGGAGACAGTATTCACC	60		

[illegible]

RESULT	4
AAT78879	ID AAT78879 standard; cDNA; 909 BP.
XX	AC AAT78879;
XX	DJ 09-OCT-1997 (first entry)
XX	DE Single chain antibody 741F8 coding sequence.
XX	KW Production; single-chain; binding protein; antibody; eukaryote; virus;
XX	KW transcription activator; promoter; expression; adenovirus; EIA; PCR;
XX	KW polymerase chain reaction; amplification; primer; herpes simplex virus;
XX	OS thymidine kinase; vector; enhancer; translation; heterologous; ss.
XX	Synthetic.
XX	Key Location/Qualifiers
FT CDS	3..755
FT FT	/tag= a
XX	/product= single chain antibody 741F8
PN US5631158-A.	
XX	PD 20-MAY-1997.
XX	PX 25-OCT-1993; 93US-0143498.
XX	PR 25-OCT-1993; 93US-0143498.
XX	PR 05-JUN-1995; 95US-0461184.
XX	PA (CREA-) CREATIVE BIOMOLECULES INC.
XX	PI Dorai H, Oppermann H;
XX	DR WPI; 1997-288577/26.
XX	RR P-PSDB; AAMW22400.
XX	PT Production of cell line for producing single-chain binding protein -
XX	PT using construct containing DNA encoding viral transcription
XX	PT activator protein
XX	PS Example 2; Column 29-32; 24pp; English.
CC	The invention relates to methods of increasing production of a
CC	single-chain binding protein, especially a single chain antibody,
CC	by generating eukaryotic cell lines containing DNA encoding either
CC	a viral transcription activator protein that acts on and stimulates
CC	a viral promoter controlling the expression of DNA encoding the

CC single-chain binding protein, such as the adenovirus Ad2 E1A protein  
CC encoded by the sequence AAT78873, or an RNA sequence able to promote  
CC translation of the RNA transcript from the heterologous gene, such as  
CC the adenoviral VAI gene (AAT78876).  
CC The sequence presented here is the coding sequence for the anti-c-erbB2  
CC single chain antibody 741F8. The sequence was constructed by amplifying  
CC the variable heavy and light chain genes which were then cloned into a  
CC pUC vector. For secretion from mammalian cells, the sequence is  
CC preceded by the signal peptide sequence from the monoclonal antibody  
CC 520C9 (AAT78880) or from Pac1 (AAT78881).  
XX  
SQ Sequence 909 BP; 229 A; 218 C; 246 G; 216 T; 0 other;

Query Match 88.3%; Score 281.8; DB 18; Length 909;  
Best Local Similarity 93.1%; Pred. No. 3e-82;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 1 GATATCGTAATGACCCAGTCTCACAATTTCATGTCACATGAGAGACATGATCAC 60  
DB 408 gatacgtcatgacccagctcctcctaaatcattcagtcagtgaggagaggtcagc 467  
QY 61 ATCACCCTCAAGGCCAGTCAGAGATGCTACTGCTGTAGCCTGGATCAACAAAACCA 120  
DB 468 atctctcgaagggcagtcagagatgtgactgtgtgactgtgtatcaaaaaacca 527  
QY 121 GGGCAATCTCTTAACCTTGATTTACTGGGCATCCACCGGACACTGGAGTCCCTGAT 180  
DB 528 gggccatctcctaaactactgatttactgacatccaccgacacacgtgagtcctgat 587  
QY 181 CGCTTCACAGGACGTGATCGTGGGACATTTACTCTCACCATCAGACGTGTGAGGCT 240  
DB 588 cgcttcacagcagtgagtcgagacattactctccacacagcagtgtagagct 647  
QY 241 GAAGACCTGGCAGCTTATTACTGTGACGACACATTTATACATCCCTGGACGTTCCGTGA 300  
DB 648 gaagaccctggcagcttacttactgacgaacattatagagtgccgtacacgttcgaggg 707  
QY 301 GGGACCAAGCTGGAGAT 317  
DB 708 gggaccaagctggagat 724

RESULT 5  
AAV21796  
ID AAV21796 standard; CDNA; 909 BP.  
AC AAV21796;  
XX  
XX 16-JUL-1998 (first entry)  
DT  
XX  
DE 741F8 anti-c-erbB-2 sfv' dimeric construct gene sequence.  
XX  
XX Antigen Imaging: single chain Fv; sfv; linker; dimeric; cancer;  
KW c-erbB-2; tumour; diagnosis; ss.  
XX  
XX Synthetic.  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT 2..755 /tag- a  
FT CDS /product- "741F8 sfv' C-terminal Gly4-Cys polypeptide"  
FT mISC\_feature 366..407 /tag- b  
FT /note- "linker sequence"  
XX  
XX US5753204-A.  
XX  
XX 19-MAY-1998.  
XX  
XX 05-JUN-1995; 95US-0461838.  
XX

PR 07-OCT-1993; 93US-0133804.  
PR 06-FEB-1992; 92US-0831967.  
PR 05-JUN-1995; 95US-0461838.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
XX Houston LL, Huston JS, Oppermann H, Ring DB;  
XX  
XX WPI: 1998-311318/27.  
XX P-PSDB: AAW53168.  
XX  
XX  
XX Imaging of antigens in vivo - using dimers of single-chain antibody  
XX Fv fragments  
XX  
XX Example 1; Columns 25-28; 30pp; English.

CC This is the nucleotide sequence of a 741F8 sfv' (single chain Fv)  
CC C-terminal Gly4-Cys construct. This was constructed by connecting the vh  
CC and vl genes with a DNA sequence encoding a 14 residue polypeptide  
CC linker. 741F8 is a monoclonal antibody useful in targeting c-erbB-2  
CC antigen. This dimeric construct can be used in the methods of invention  
CC of imaging a preselected antigen expressed in a mammal. The methods are  
CC used in magnetic resonance imaging of c-erbB-2 or related antigens in  
CC cancer diagnosis. The biosynthetic constructs have enhanced properties  
CC as in vivo targeting agents in comparison with intact monoclonal  
CC antibodies or their Fab fragments. The dimeric constructs permit the  
CC in vivo targeting of an epitope on an antigen with greater apparent  
CC avidity, including greater tumour specificity, tumour localisation and  
CC tumour retention properties than that of the Fab fragment having the same  
CC CDS as the construct.

SQ Sequence 909 BP; 229 A; 218 C; 246 G; 216 T; 0 other;

Query Match 88.3%; Score 281.8; DB 19; Length 909;  
Best Local Similarity 93.1%; Pred. No. 3e-82;  
Matches 295; Conservative 0; Mismatches 22; Indels 0; Gaps 0;  
QY 1 GATATCGTAATGACCCAGTCTCACAATTTCATGTCACATGAGAGACATGATCAC 60  
DB 408 gatacgtcatgacccagctcctcctaaatcattcagtcagtgaggagaggtcagc 467  
QY 61 ATCACCCTCAAGGCCAGTCAGAGATGCTACTGCTGTAGCCTGGATCAACAAAACCA 120  
DB 468 atctctcgaagggcagtcagagatgtgactgtgtgactgtgtatcaaaaaacca 527  
QY 121 GGGCAATCTCTTAACCTTGATTTACTGGGCATCCACCGGACACTGGAGTCCCTGAT 180  
DB 528 gggccatctcctaaactactgatttactgacatccaccgacacacgtgagtcctgat 587  
QY 181 CGCTTCACAGGACGTGATCGTGGGACATTTACTCTCACCATCAGACGTGTGAGGCT 240  
DB 588 cgcttcacagcagtgagtcgagacattactctccacacagcagtgtagagct 647  
QY 241 GAAGACCTGGCAGCTTATTACTGTGACGACACATTTATACATCCCTGGACGTTCCGTGA 300  
DB 648 gaagaccctggcagcttacttactgacgaacattatagagtgccgtacacgttcgaggg 707  
QY 301 GGGACCAAGCTGGAGAT 317  
DB 708 gggaccaagctggagat 724

RESULT 6  
AAV22598  
ID AAV22598 standard; CDNA; 909 BP.  
AC AAV22598;  
XX  
XX 13-JUL-1998 (first entry)  
DT  
XX  
DE DNA encoding single chain binding site molecule (sfv) 741F8.  
XX







Query Match	85.18;	Score 271.6;	DB 16;	Length 360;
Best Local Similarity	90.98;	Pred. No. 4.6e-79;		
Matches 289; Conservative	0;	Mismatches 29;	Indels 0;	Gaps 0

RESULT	10
AAQ90424	
ID	AAQ90424 standard; DNA; 321 BP.

DE DNA encoding immunoglobulin light chain of anti-idiotypic antibody  
DE against human anticancer antibody.

KW Antibody; cancer; CDR; heavy chain; light chain; immunoglobulin;  
KW complementarity determining region; ss.

PN JP07101999-A.

PD 18-APR-1995.

PF 06-OCT-1993; 93JP-0272950.

PR 06-OCT-1993; 93JP-0272950.

PA (HAGI/) HAGIWARA Y.

DR WPI; 1995-182987/24.

XX

PT Novel anti-idiotypic antibody against an human anticancer monoclonal antibody - and DNA sequences encoding the antibody, useful in pharmacology, medicine and biochemical fields.

PS Claim 22; Page 5; 28pp; Japanese.

Query Match	84.1%	Score 268.4	DB 16	Length 321
Best Local Similarity	90.3%	Pred. No. 5e-78		
Matches 287, Conservative	0	Mismatches 31	Indels 0	Gaps 0

[illegible]

RESULT	ID
11	AAV39345
	AAV39345 standard; CDNA; 394 BP.

XX 21-SEP-1998 (first entry)  
DT

DE Mouse anti-HM1.24 antibody light chain variable region encoding cDNA.

KM Mouse; human; humanised; anti-HM1.24 antibody; myeloma; FR; CDR; KW framework region; complementarity determining region; antigenicity; ss

Mus sp.

...	FH	Key	Location/Qualifiers
-----	----	-----	---------------------

FT                    / \*tag= a

XX

XX :

XX

XX

PR 04-OCT-1996; 96JP-0264756.  
XX  
PA (CHUS ) CHUGAI SEIYAKU KK.

PI Koishihara Y, Kosaka M, Ohtomo T, Ono K, Tsuchiya M,

XX WP1: 1998-286421/25.  
DR P-PSDB: AAM62187.  
XX  
PT Humanised anti-HM1.24 antibody - for treatment of myeloma  
XX  
PS  
PS Claim 2; Page 100-101; 210pp; Japanese.  
XX  
XX A humanised anti-HM1.24 antibody has been developed which comprises  
CC human L and H chain C regions, and L and/or H chain V regions  
CC containing material originating in mouse anti-HM1.24 antibody. The V  
CC regions contain framework (FR) regions of human origin and  
CC complementarily determining regions (CDR) of mouse origin, leading to  
CC a reshaped humanised antibody. The C regions are human Ck (L-chain) and  
CC human C gamma (especially C gamma 1) (H-chain). The FR regions of the  
CC L chain V region are derived from human subtype HSG1 (e.g. from human  
CC antibody RE1) and the FR regions of the H chain V region are derived  
CC from human subtype HSG1 (e.g. FR-3 from human antibody HG3 and FR4  
CC from human antibody JH6). The present sequence encodes a mouse L  
CC chain V region from the present invention. The antibodies are used for  
CC the treatment of myeloma, especially by injection, intravenously,  
CC intramuscularly or subcutaneously. The antibodies are used at 0.01-1000  
CC (especially 5-100) mg/kg body weight. The humanised antibody has low  
CC antigenicity and is therefore effective therapeutically in humans.  
XX  
XX Sequence 394 BP; 102 A; 93 C; 95 G; 104 T; 0 other;  
XX

XX Mus.sp.  
OS JP07101999-A.  
XX  
XX 18-APR-1995.  
PD  
XX 06-OCT-1993; 93JP-0272950.  
PF  
XX 06-OCT-1993; 93JP-0272950.  
PR  
XX (HAGI/) HAGIWARA Y.  
PA WPI; 1995-182987/24.  
DR P-PSDB; AAR74958.  
XX  
XX Novel anti-idiotype antibody against an human anticancer monoclonal  
PT antibody - and DNA sequences encoding the antibody, useful in  
PT pharmacology, medicine and biochemical fields.  
XX  
XX Claim 21, Page 5; 28pp; Japanese.  
PS  
XX AA090420-Q090424 are DNA molecules encoding possible light chains of a  
CC new anti-idiotype antibody against a human anticancer monoclonal  
CC antibody. This antibody contains in its heavy chain 3 complementarily  
CC determining regions CDRI (AAR74929-R74931), CDRI2 (AAR74932-R74935) and  
CC CDRI3 (AAR74936-R74939), this is also true of the light chain which has  
CC its own CDRI (AAR74944-R74946 and AAR85774), CDRI2 (AAR74947-R74949) and  
CC CDRI3 (AAR74950-R74954) The antibody and DNA encoding it are useful in  
CC pharmacology, medical and biochemical fields.  
XX  
XX Sequence 321 BP; 83 A; 84 C; 75 G; 79 T; 0 other;  
SO

Query Match 81.9%; Score 261.2; DB 16; Length 321;  
Best Local Similarity 89.5%; Pred. No. 1.1e-75;  
Matches 281; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 GATTCGTAATGACCCAGTCTCACAAATTCATGTCACATGAGGAGACGATGATCAC 60  
DB 1 gacatctgtatgaccagctccacaatctcatgcccacatcagtagagacagagctcagt 60  
QY 61 ATCACTGCAAGGCGCAGTGCAGATGATGACTACTGCTAGCCCTGTATCAACAAACCA 120  
DB 61 atcacctgcgaagcgccagctcagatgtagaactctgtagctgtatcaacgagaacca 120  
QY 121 GGGCAATCTCTTAACCTTGTGATTTACTGGGATCCACCGCGACACTGGAGTCCCTGAT 180  
DB 121 ggaacaatctcttaaaccttactgcttactcgcacatcctacggtacactgagatccctgat 180  
QY 181 CGGTTACAGGAGTGGATGTCGAGATTTACTCTCAGCAGCAGCAGGATGTCAGGCT 240  
DB 181 cactctcgtgcagctgagatcgggagcgatctcacttcacacacagcgagctcagcgt 240  
QY 241 GAAGACCTGGCAGCTTTATTTACTGTCAGCAACATTATACACTCCGTTGGTGGGA 300  
DB 241 gaagacctggcagcttattactctgctcagcaaatattatagctcctcctccacgcttgagct 300  
QY 301 GGGACACAGCTGGA 314  
DB 301 gggacacaagctgga 314

RESULT 14  
ID ABA94232 standard; DNA; 303 BP.  
XX ABA94232:  
XX  
XX 13-MAR-2002 (first entry)  
DT  
XX Mouse 04 antibody kappa light chain 2 variable region DNA.  
XX

KW Neuromodulatory; central nervous system; CNS; SH1GM22; LYM 22; AKR4;  
KW ebvH1GM MS119D10; ebv H1GM CB2b8; CB21E12; CB21E7; MS119E5; v1rucide;  
KW antiparkinsonian; neuroprotective; nootropic; vulnerrary; mouse; ds.  
XX  
XX Mus.sp.  
OS  
XX  
XX Key Location/Qualifiers  
FH 1..303  
FT CDS /\*tag= a  
XX  
XX WO200185797-A1.  
XX  
XX 15-NOV-2001.  
PD  
XX 30-MAY-2000; 2000WO-US14902.  
PF  
XX 10-MAY-2000; 2000US-0568351.  
PR  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION RES.  
XX  
XX Rodriguez M, Miller DJ, Pease LR;  
PI WPI; 2002-066596/09.  
DR P-PSDB; ABB07185.  
XX  
XX Novel neuromodulatory agent (a human IGM monoclonal antibody),  
PT promoting neurite outgrowth, regeneration, remyelination and  
PT neuroprotection in central nervous system, useful to treat  
PT post-infectious encephalomyelitis -  
XX  
XX Disclosure; Fig 44; 219pp; English.  
PS

CC The invention provides a neuromodulatory agent (1) capable of promoting  
CC neurite outgrowth, regeneration, remyelination and neuroprotection in  
CC central nervous system (CNS). (1) is capable of inducing remyelination,  
CC promoting cellular proliferation of glial cells, and promoting Ca2+  
CC signaling with oligodendrocytes. An humanised antibody to (1) can be  
CC selected from antibody SH1GM22 (LYM 22), ebvH1GM MS119D10, ebv H1GM  
CC CB2b8, AKR4, CB21E12, CB21E7 or MS119E5. (1) is useful for stimulating  
CC remyelination of CNS axons, stimulating proliferation of glial cells in  
CC CNS axons, or treating demyelinating disease of CNS in a mammal in need  
CC of such therapy. (1) is capable of binding to structures and cells within  
CC CNS. (1) is preferably useful for treating a demyelinating disease of CNS  
CC of a mouse infected with strain DA of Theiler's murine encephalomyelitis  
CC (TMEV) or for treating a human being having multiple sclerosis, or a  
CC human or domestic animal with a viral demyelinating disease, or a post-  
CC neural disease of CNS. (1) is also useful for an in vitro method of  
CC stimulating the proliferation of glial cells from mixed cell culture.  
CC (1) is also useful for stimulating remyelination of CNS axons. The  
CC antibodies are useful for preventing infection by a bacterium, virus or  
CC like pathogen that causes demyelination or other neurodegenerative  
CC condition in a subject. Methods where (1) is administered to a patient  
CC are useful for treating multiple sclerosis, Parkinson's disease,  
CC Alzheimer's disease, amyotrophic lateral sclerosis (ALS), a viral  
CC demyelinating disease, CNS diseases, and other conditions in the CNS  
CC where nerves are damaged as by trauma. The present sequence represents  
CC the mouse 04 antibody kappa light chain 2 variable region DNA.  
XX  
XX Sequence 303 BP; 76 A; 80 C; 72 G; 75 T; 0 other;  
SO

Query Match 81.4%; Score 259.8; DB 24; Length 303;  
Best Local Similarity 91.1%; Pred. No. 3.2e-75;  
Matches 276; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 GATATCGTAATGACCCAGTCTCACAAATTCATGTCACATGAGGAGACGATGATCAC 60  
DB 1 gacatctgtatgaccagctccacaatctcatgcccacatcagtagagacagagctcagt 60  
QY 61 ATCACTGCAAGGCGCAGTGCAGATGATGACTACTGCTAGCCCTGTATCAACAAACCA 120  
DB 61 atcacctgcgaagcgccagctcagatgtagaactctgtagctgtatcaacgagaacca 120



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 05:38:45 ; Search time 2643.23 Seconds  
(without alignments)  
1628.889 Million cell updates/sec

Title: US-09-824-286-5

Perfect score: 319

Sequence: 1 GATATCGTATGACCCAGTTC.....AGGACCAAGCTGAGATCT 319

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estbda:\*  
2: em\_estbhm:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281.8	88.3	974	10	BF578595 602093306
2	276.2	86.6	960	10	BF584737 602098485
3	251.8	78.9	985	10	BF581877 602098343
4	249.4	78.2	841	10	BF581734 602095514
5	236.6	74.2	887	10	BF584391 602096577
6	235.6	73.9	680	10	BE376043 601229480
7	232.8	73.0	594	10	BE309592 601094848
8	231.4	72.5	345	9	AA710291 v153a04.r
9	231.2	72.5	673	12	BH065089 RPCI-24-3
10	230.6	72.3	950	10	BF781701 602104224
11	224.6	70.4	958	10	BF135931 601781261
12	221.6	69.5	514	12	BH104821 RPCI-24-2
13	220	69.0	881	10	BE913824 601669294
14	214	67.1	899	10	BG973020 602838676
15	210.4	66.0	883	10	BF580610 602093730
16	196	61.4	471	9	AI233978 EST230666
17	195.6	61.3	864	10	BH11610 602895973

18	193.4	60.6	871	10	BG962566
19	183.8	57.6	867	10	BG754732
20	183.6	57.6	891	10	BG540787
21	183.6	57.6	992	10	BP976253
22	182	57.1	826	10	BG684416
23	182	57.1	957	10	BG341934
24	182	57.1	959	10	BG341941
25	180.4	56.6	417	9	AA406228
26	180.4	56.6	422	9	AA407904
27	180.4	56.6	714	10	BG757897
28	180.4	56.6	750	10	BG755394
29	180.4	56.6	1130	10	BF662874
30	180.4	56.6	1198	10	BF663671
31	179	56.1	711	10	BG546866
32	179	56.1	867	10	BG539961
33	178.8	56.1	431	9	AA406886
34	178.8	56.1	471	9	AA406294
35	178.8	56.1	886	10	BG756818
36	177.8	55.7	858	10	BG756326
37	177.4	55.6	927	10	BG536845
38	177.2	55.5	426	9	AA405900
39	177.2	55.5	493	9	AA405753
40	176.4	55.3	712	10	BG538665
41	176.2	55.2	813	10	BG431143
42	175.8	55.1	431	10	B1305095
43	175.8	55.1	721	10	BG686441
44	175.8	55.1	727	10	BG684242
45	175.6	55.0	444	9	AV686883

#### ALIGNMENTS

RESULT 1  
LOCUS BF578595 974 bp mRNA linear EST 12-DEC-2000  
DEFINITION 602093306F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4207632 5',  
mRNA sequence.  
ACCESSION BF578595  
VERSION BF578595.1 GI:11652410  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 974)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
plate: LMA9770 row: 1 column: 01  
High quality sequence stop: 711.  
Location/Qualifiers  
1..974

#### FEATURES

source

1..974  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4207632"  
/clone\_id="NCI\_CGAP\_Co24"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: colon; Vector: PCMV-SPOrt6; Site: 1; Noti;  
Site-2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 281 a 262 c 232 g 199 t

ORIGIN

Query Match	88.3%;	Score 281.8;	DB 10;	Length 974;
Best Local Similarity	93.1%;	Pred. No. 1.1e-77;		
Matches 295; Conservative	0;	Mismatches 22;	Indels 0;	Gaps 0;

QY	1	GATATCGTAAAGACCCAGCTCACAATAATTCATATGTCACATCGATGAGGAAACAGTATCACC	60
Db	80	GACATTTGTATGACCCAGCTGCACAAATTCATATGTCACATCGATGAGGAAACAGGCTCACC	139
QY	61	ATACGCTCAAGGCGCAGTCAGATGTGACTACTGCTGTATAGCTGGTATATACAAAAACCA	120
Db	140	ATACGCTCAAGGCGCAGTCAGATGTGACTACTGCTGTATATACGTTATACAAAAACCC	199
QY	121	GGGCAATCTCCTAAACTTCTGATTACTAGTGGGCGATCCACCCGGCAGACTGGAATCCTCTGAT	180
Db	200	GGGCTATCTCCTAAACTATTTGATTACTAGGGCATCCACCCGGCAGACTGGAATCCTCTGAT	259
QY	181	CGCTTCACAGGCAGTCGATCTTGGGACAGATTATCTTCACCATCAGCAGTGTGCAGGCT	240
Db	260	CGCTTCACAGGCAGTGGGCTCTGGGACAGATTATCTTCACCATCAGCAGTGTGCAGGCT	319
QY	241	GAAGACCTGGGCACTTTATTACTCTGAGCAACAATATATCACTCCGGAGAGCTTGGGTGA	300
Db	320	GAAGACCTGGGCACTTTATTACTCTGAGCAACAATATATCACTCCGGAGAGCTTGGGTGCT	379
QY	301	GGGACCAAGCTGGAGAT	317
Db	380	GGGACCAAGCTGGAGCT	396

RESULT	2
LOCUS	Bf584737
DEFINITION	Bf584737 960 bp mRNA EST 12-DEC-2001 6020964485F1 NCL_CGAP_CO24 Mus musculus cDNA clone IMAGE:4218357 5'
ACCESSION	Bf584737
VERSION	Bf584737
KEYWORDS	Bf584737.1 GI:11658455
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 960)
AUTHORS	NIH-MGC ( <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> ). National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

BASE COUNT	260 a	251 c	231 g	218 t
ORIGIN				

Query Match	86.68;	Score 276.2;	DB 10;	Length 960;
Best Local Similarity	94.08;	Pred. No. 6e-76;		
Matches 298; Conservative	0;	Mismatches 18;	Indels 1;	Gaps 1;

Qy	1	GATTTTCGTAATGAGCCAGCTTCACAAATTCATGTCCATCATGTAAGGAGACAGTATCCAC	60
Db	53	GACATTTGTGATGAGCCAGCTTCACAAATTCATGTCCATCATGTAAGGAGACAGGCTCAGC	112
Qy	61	ATCACCTGCAGAGCCAGTCAGAGATGTGACTCTCTGCTGATAGCTGGTATCAACAAANAACA	120
Db	113	ATCACCTGCAGAGCCAGTCAGAGATGTGACTCTCTGCTGATAGCTGGTATCAACAAANAACA	172
Qy	121	GGGCAATCTCCTAAACTTCTGATTTACHTGGGCATCCACCCGGCACACTGGAGTCCCTGAT	180
Db	173	-GGCTATCTCTAAACCTGATTTACTTACGGGCATCCACCCGGCACACTGGAGTCCCTGAT	231
Qy	181	CGCTTCCACGACAGTGGATGTGGACACATATATCTCCACATACAGAGTGTGAGGCT	240
Db	232	CGCTTCCACGACAGTGGATGTGGACACATATATCTCCACATACAGAGTGTGAGGCT	291
Qy	241	GAAGACCTGGGCACTTTATCTGTCAGCAACAATTTATTCACCTCCGTGGACGTTTGGTGA	300
Db	292	GAAGACCTGGGCACTTTATCTGTCAGCAACAATTTATTCACCTCCGTGGTGGT	351
Qy	301	GGGACCAAGCTGGAGAT	317
Db	352	GGGACCAAGCTGGAGCT	368

RESULT	3
RF581877	
LOCUS	
DEFINITION	985 bp mRNA linear EST 12-DEC-2000 60209934.F1 NCL_CGAP_CO24 Mus musculus cDNA clone IMAGE:4219018 5' ,
ACCESSION	BF581877
VERSION	BF581877
KEYWORDS	BF581877.1 GI:11655693
SOURCE	EST.
ORGANISM	house mouse.  Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 985)
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

BASE COUNT	291 a	252 c	220 g	222
ORIGIN				





Query Match	73.9%;	Score 235.6;	DB 10;	Length 680;
Best Local Similarity	84.4%;	Pred. No. 3e-63;		
Matches 265; Conservative	0;	Mismatches 49;	Indels 0;	Gaps 0;

[illegible]

RESULT 7  
BE309592

\*\*\*\*\*  
FOA b-  
-----  
DATE OF COM 0000

LOCUS	BE309592	594	bp	mRNA	linear	EST:26-Oct-2000
DEFINITION	BC109484ef1 NC1.CGAP_Mam5 Mus musculus cDNA clone IMAGE:3489635 5'					
ACCESSION	BE309592					
VERSION	BE309592.1					
KEYWORDS	EST.					
ORGANISM	house mouse.					
SOURCE	Mus musculus.					

REFERENCE AUTHORS TITLE JOURNAL COMMENT
<p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Mus. 1 (bases 1 to 594)</p> <p>NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a>.</p> <p>National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)</p> <p>Contact: Robert Strausberg, Ph.D.</p>

Email: [cgdp@fremail.hhi.nyu.gov](mailto:cgdp@fremail.hhi.nyu.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LHAM530 row: 0 column: 12  
High quality sequence stop: 591  
[cgdp@fremail.hhi.nyu.gov](mailto:cgdp@fremail.hhi.nyu.gov)

```

FEATURES
Source
Location/Qualifiers
1. .594
/Organism="Mus musculus"
/strain="c57/bl6"

```

```

/bed_xref="57720"
/bed_xref="taxon:10090"
/clone="IMAGE:3489635"
/clone_1b="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by life technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

```

BASE COUNT ORIGIN	158 a	149 c	143 g	144 e
----------------------	-------	-------	-------	-------



Resources (<http://www.chori.org/Bacpac/orderinfoframe.htm>). BAC end  
 Page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 370 Row: 0 Column: 10  
 Seq primer: T7  
 Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .673

BASE COUNT	185 a	158 c	134 g	196 t
ORIGIN				

	Query Match	72.5%	Score 231.2	DB 12	Length 673	
	Best Local Similarity	88.4%	Pred. No. 7.3e-62			
	Matches	251	Conservative	0	Mismatches	33
					Indels	0
					Gaps	0
QY	1	GATATCGTAATGACCCAGCTCTCACAATTCATGTCCACATCGATGAGAGACGATATCACC	60			
DB	288	GACATTTGGAATGACCCAGCTCTCAAAATTCATGCTCACATCGATGAGACAGGGGCACG	347			
QY	61	ATCACCTCTCAAGGCGCAGTCAGCATGTGCACTACTGTGTCGTGATCAACAAAACCA	120			
DB	348	ATCACCTGCAAGGCCAGTCAGATGTGGGTACTGCTGTAGCCTGTGATTAACAGAAAACA	407			
QY	121	GGGCAATCTCTTAACCTTCTGATTTTACTTGGGGATATCCACCCGGCAGACACTGGAATCCTGAT	180			
DB	408	GGACATATCTCTAAACCTACTGATTTTACTGCGGATCCAAATCGGTATCACTGTGAATCCCTGAT	467			
QY	181	CGCTTCACAGGCGAGTGATCTGGGACAGATTTTACTCTACCATCAGCAGTGTGAGGCT	240			
DB	468	CGCTTCACAGGCGAGTGATCTGGGACAGATTTTACTCTACCATCAGCAGATTTGAGTCT	527			
QY	241	GAAGCCTGGCACTTATTACTGTCAGCAACATTTATATCACC	284			
DB	528	GAAAGCCTGGCAGATTAATTTCTGCCAGCAATATACAGATATCC	571			

RESULT	10
LOCUS	Bf781701
DEFINITION	Bf781701 950 bp mRNA linear EST 12-JAN-2001 602104224F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4222684
ACCESSION	5', mRNA sequence.
VERSION	Bf781701
KEYWORDS	Bf781701.1 GI:12086737
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 950)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	Email: gcgaps-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>

```

Plate: L1AM9809 row: 0 column: 05
High quality sequence stop: 617.
Location/Qualifiers
1..950
FEATURES
source
```

BASE COUNT	252 a	226 c	247 g	224 t	1 others
ORIGIN					
Query Match	72.3%	Score 230.6	DB: 10	Length 950	
Best Local Similarity	83.0%	Pred. NO. 1.3e-61			
Matches 263: Conservative	0	Mismatches 34	Indels 0	Gaps 0	

QY	1	GATATCGTAATGAGCCACCTCTCACAAATTCATGTCACATCGATGAGAGACAGTATCAC	60
Db	80	GACATTTGTGCTGACCCACTGCTCAAAAATTCATGTCACATCGATGAGAGACAGGGTCCAC	139
QY	61	ATCACCTCGAAGGCCAGTCAGAGATGTGACTACTGCTGTAGCCTGGTATTCACAAAAACCA	120
Db	140	ATCACCTCGAAGGCCAGTCAGAGATTTTGTGATCTGCTGTAGCCTGGTATTCACAAAAACCA	199
QY	121	GGGCAATCTCCTAACTTCTGATTTACTTGTGGGCATCCACCCGGCACACTGSAATCCCTGAT	180
Db	200	GGGCGAGGCTCCTAAACCACTAATTTATTTATGTGATCCAAACCCGACACTGSAATCCCTGAT	259
QY	181	CGCTTCACAGGACGATGAGATCTGGGACATATTTACTCTCACATPAGACAGTGTGAGGCT	240
Db	260	CGCTTCACAGGACGATGAGATCTGGGACATTTTCACTCTCACATPAGACAGTGTGAGGCT	319
QY	241	GAAGACCTGGCATTATTTACTGTGAGCAGACATTTATCACTCCGTGACAGTTCCGTGGA	300
Db	320	GAAGACCTGGCAGATATTTCTGTGTGAGCAGATGTGGAATTTATCTTCACAGTTCCGAGGG	379
QY	301	GGGACCAAGCTGTGAGAT	317
Db	380	GGGACCAAGTTGGAAT	396

RESULT	11				
LOCUS	BF135931				
DEFINITION	BF135931	918 bp	mRNA	linear	EST 24-OCT-2000
ACCESSION	601781261	NCI_CGAP_Lu30	Mus musculus	CDNA clone	IMAGE:4009430 5',
VERSION	BF135931				
KEYWORDS	BF135931.1	GI:10974971			
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 918)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://Image.lnl.gov">http://Image.lnl.gov</a> Plate: L1AM9245 row: 1 column: 15				





CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>

Plate: LLM9771 row: 1 column: 07  
 High quality sequence stop: 639.

## FEATURES

## SOURCE

Location/Qualifiers  
 1. 883  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4208022"  
 /clone\_lib="NCI CGAP Co24"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: colon; Vector: PCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.6 kb. Constructed by Life  
 Technologies. Note: this is a NCI CGAP library."

## BASE COUNT

233 a 219 c 229 g 202 t

## ORIGIN

## Query Match

Best Local Similarity 66.0%; Score 210.4; DB 10; Length 883;

Matches 263; Conservative 83.2%; Pred. No. 2.8e-55;

Mismatches 0; Mismatches 51; Indels 2; Gaps 2;

```

OY 3 TATGTAATGACCCAGTCTCACAATTCATGTCACATCAGTAGAGACAGTATCACCAT 62
    ||| | ||||| ||| ||||| ||| ||||| ||||| ||||| |||||
Db 74 TATTTGATGACCCAGTCTCACAATTCCTG-CTGTATCAGCAGAGAGGTTACCAT 132
OY 63 CACCTGCAAGGCCAGTCTGAGTGTACTGCTGTAGCTGGTATCAACAAAACCAAG 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133 AACCTGCAAGGCCAGTCTGAGTGTATGATGTAG-CTGTACCAACAGAGTCAAG 191
OY 123 GCAATCTCCTAACTTCTGATTTACTGGGATCCACCCGGCACACTGGAGTCCCTGATCG 182
    ||| ||||| ||||| ||||| ||||| ||| ||||| ||||| |||||
Db 192 GCAGTCTCCTAAAGCTGATATATCTATCATCCAAATCGTATACTGGAGTCCCTGATCG 251
OY 183 CTTGACAGGAGTGGATGGAGAGATATATCTCACCATCAGCAGTGTGCAGGCTGA 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 CTTGACTGGCAGTGGACATGGAGGATTTCACTTTCACCATCAGCAGTGTGCAGGCTGA 311
OY 243 AGACCTGGCAGTGTATTTACTGTCAGCAACATTATTCACCTCCGTGAGCTTGGTGAGG 302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 AGACCTGGCAGTGTATTTCTGTGACAGAGATTATAGTCTCTCGAGAGTTGGTGAGG 371
OY 303 GACCAAGCTGGAGATC 318
    ||||| ||||| |||
Db 372 CACCAAGCTGGAAATC 387

```

Search completed: August 20, 2002, 08:55:02  
 Job time: 11777 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 09:25:29 ; Search time 1815.27 Seconds

(without alignments)  
3873.427 Million cell updates/sec

Title: US-09-824-286-6

Perfect score: 336

Sequence: 1 CTCGACGAGTCACGACCTGCG.....GCCACAGGACACGACGACAC 336

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: \*  
1: gb\_da: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_jnu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hlg\_hum: \*  
31: em\_hlg\_inv: \*  
32: em\_hlg\_other: \*  
33: em\_higo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match Length	DB ID	Description
------------	-------	-------	--------------	-------	-------------

1	302.4	90.0	759	6	AX256292	AX256292 Sequence
2	292.8	87.1	354	10	AF178588	AF178588 Mus muscu
3	283.8	84.5	354	10	MUSAB	L48671 Mus musculu
4	283.4	84.3	363	6	AX084361	AX084361 Sequence
5	283.4	84.3	363	6	AX286236	AX286236 Sequence
6	283	84.2	354	10	MUSIGHARI	M36228 Mouse Ig he
7	279.8	83.3	357	6	AX256252	AX256252 Sequence
8	277	82.4	360	10	MUSIGHAMX	AX256252 Sequence
9	276.8	82.4	350	10	S76531	M36217 Mouse Ig he
10	275.4	82.0	345	10	MUSIGMO520	S76531 Ig VH-antli-
11	274.4	81.7	345	10	MMU60456	M76414 Mouse IgM c
12	274.2	81.6	342	10	AF006578	U60456 Mus musculu
13	273.6	81.4	351	6	AR083821	AF006578 Mus muscu
14	273.6	81.4	351	10	MMASMSIH	AR083821 Sequence
15	273.6	81.4	420	6	AR169919	X75099 M.musculu
16	273.2	81.3	331	10	AF072800	AR169919 Sequence
17	273	81.2	333	10	MUSIGKCLJ	AF072800 Mus muscu
18	272.2	81.0	363	10	MMACHVR8	M15233 Mouse IgM c
19	271	80.7	318	10	MMHGVF	X90883 M.musculu
20	271	80.7	1431	6	AX055011	X58646 Mouse reatr
21	271	80.7	1431	6	AX055012	AX055011 Sequence
22	270.8	80.6	351	10	AF110502	AX055012 Sequence
23	270.4	80.5	337	10	MUSIGHBL	AF110502 Mus muscu
24	269.8	80.3	304	6	AR083817	K00721 Mouse Ig ac
25	269.6	80.2	337	10	MUSIGHBJ	AR083817 Sequence
26	269.6	80.2	340	10	MUSIGHBM	K00719 Mouse Ig ac
27	269.6	80.2	341	10	MUSIGHBK	K00722 Mouse Ig ac
28	269.6	80.2	346	10	MUSIGHBO	K00720 Mouse Ig ac
29	269.6	80.2	348	10	MUSIGHBN	K00724 Mouse Ig ac
30	269.6	80.2	376	10	MMAPHOXH	K00723 Mouse Ig ac
31	269.6	80.2	401	10	MMU53526	X15471 Mouse partI
32	269.6	80.2	668	6	A25750	U53526 Mus musculu
33	269.6	80.2	669	6	AR051417	A25750 Variable re
34	269.6	80.2	708	6	A25744	AR051417 Sequence
35	269.6	80.2	708	6	AR051414	A25744 Variable re
36	269.4	80.2	321	10	MMHGVF	AR051414 Sequence
37	269	80.1	387	10	AF276284	X58649 Mouse reatr
38	268.2	79.8	304	10	AF118978	AF276284 Mus muscu
39	268	79.8	339	10	S82436	AF118978 Mus muscu
40	268	79.8	342	10	MUSIGHBR	S82436 Ig VH-H2 Im
41	267.8	79.7	342	10	MUSIGHBP	K00727 Mouse Ig ac
42	267.8	79.7	355	10	MMU55482	K00725 Mouse Ig ac
43	267.6	79.6	291	6	AR083822	U55482 Mus musculu
44	267.6	79.6	291	6	AR100553	AR083822 Sequence
45	267.6	79.6	291	10	MMVM13	AR100553 Sequence

## ALIGNMENTS

RESULT 1  
AX256292  
LOCUS AX256292 759 bp DNA  
DEFINITION Sequence 62 from Patent WO0171005.  
ACCESSION AX256292  
VERSION AX256292.1 GI:16075164  
SOURCE  
KEYWORDS  
ORGANISM  
synthetic construct.  
artificial sequence.  
REFERENCE  
1 (bases 1 to 759)  
AUTHORS  
Kufner, P., Rietmueller, G., Luterbuese, R., Borschert, K.,  
Kischel, R., Mayer, M. and Hofmeister, R.  
TITLE  
Multifunctional polypeptides comprising a binding site to an  
epitope of the nkx2d receptor complex  
JOURNAL  
Patent: WO 0171005-A 62 27-SEP-2001;  
Kufner, Peter (DE)  
FEATURES  
Location/Qualifiers  
1..759  
BASE COUNT 188 a 186 c 201 g 184 t  
ORIGIN

```

/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/cell_line="18-1"
/cell_type="hybridoma"
<1..>354
CDS
/note="cross-reactive antibody"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="AAD5344.1"
/db_xref="GI:5853166"
/translation="DVQLKSGPGIIVAPSGLSITCTVSGFSLTSYGVHWYKPPGKG
LEMLIVAGSGTNYNSALMSRLSISKMSKQYFLKMSLQTDITMYCARQYGRSR
GAMDWGGGTSTVYS"
BASE COUNT
90 a      89 c      94 g      81 t
ORIGIN

```

	Query Match	84.5%	Score 283.8	DB 10	Length 354	
	Best Local Similarity	90.4%	Pred. No. 9.e-75			
	Matches 303	Conservative	0	Mismatches 32	Indels	Gaps
Qy	1	CTCAGAGACGACGACTGGCGTGGGGCCCTCACAGAGCCTGTGCATCACTTCACAT	60			
Db	10	CTCAGAGAGTCAGAGACTGTCTCGTGGGGCCCTCACAGAGCCTGTGCATCACTTCGACT	69			
Qy	61	GTCCTCGGTTTCATTAAACGACGTAATGGTGTACACTGGGTTGCCAGAGCCTCCAGGAAGA	120			
Db	70	GTCCTCGGTTTCATTAAACGACGTAATGGTGTACACTGGGTTGCCAGAGCCTCCAGGCAGA	129			

Db	190	ATGTCACAGTCGAGCATCGACAAAGACAACATCCAGAGCCAAAGTTTCTTTAAATAATGAAC	249
Qy	241	AGTCGCAAACTGATGACACAGCCATCTACTACTGTGCGCAGAGAGGGTTC-----T	291
Db	250	AGTCGACAGACTGATGACACAGCCATCTACTACTGTGCGCAGAGATCATCTACTATGATT	309
Qy	292	ACGGTAGATTCTATGAGTACACTGCGGGGCCAAGGAGGACGAGTCACC	336
Db	310	ACGGCTATGCTATGACTACTGCGGGGCCAAGGAGGACCGAGTCACC	354
RESULT	5		
LOCUS	AX286236	363 bp	DNA linear PAT 20-NOV-2001
DEFINITION	Sequence 2 from Patent EP1076089.		
ACCESSION	AX286236		
VERSION	AX286236.1	GI:17046098	
KEYWORDS			
SOURCE	Mus sp.		
ORGANISM	Mus sp.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Deckmyn,H. and Cauwenberghs,N.		
TITLE	Cell lines, ligands and antibody fragments for use in pharmaceutical compositions for preventing and treating haemostasis disorders		
JOURNAL	Patent: EP 1076089-A 2 14-FEB-2001;		
FEATURES	K.U. LEUVEN RESEARCH & DEVELOPMENT (BE)		
source	1..363		
	/organism="Mus sp."		
	/db_xref="taxon:10095"		
primer_bind	1..24		
	/note="Vhback primer"		
V.region	1..363		
misc_feature	90..105		
	/note="complementary determining region one"		
misc_feature	147..195		
	/note="complementary determining region two"		
misc_feature	291..330		
	/note="complementary determining region three"		
primer_bind	330..360		
	/note="Vhfor primer"		
BASE COUNT	93 a 91 c 95 g 84 t		
ORIGIN			
Query Match	84.3%;	Score 283.4;	DB 6; Length 363;
Best Local Similarity	91.3%;	Pred. No.1.3e-74;	
Matches 315; Conservative	0;	Mismatches 21;	Indels 9; Gaps 1;
Qy	1	CTGCAGGAGTCAAGCACTCGGCGCTGTGGGGCCCTCAACAGAGCCGTGTCATCATCTGCACT	60
Db	10	CTGCAGGAGTCTGGCACTGCGCTGGTGGCGCCCTCACAGGCTCTCCATCTGCTGCACT	69
Qy	61	GTCCTGGGTTTCTATTAAACAGCATATGGTGTACACTGGGTTCCGACAGCTCCAGAAAG	120
Db	70	GTCCTGGGATTTCATTAAACAGATATGGTGTACACTGGGTTCCGACAGCTCCAGAAAG	129
Qy	121	GGTCTGAGAGTGGCTGGAGTCAATTTGGCTGGTGAAGACACAATTAATTAATTCGGCTCTC	180
Db	130	GGTCTGAGAGTGGCTGGAGTCAATTAATGAGTGGTGAAGACACAATTAATTAATTCGGCTCTC	189
Qy	181	ATGTCCACACAGCAATCATCAACGAGACAATTCACAGAGCCAAATTTCTTAAATAATGAAC	240
Db	190	ATGTCCACACAGTATGATCATCAACGAGACAATTCACAGAGCCAAATTTCTTAAATAATGAAC	249
Qy	241	AGTCGCAAACTGATGACACAGCCATCTACTACTGTGCGCAGAGAGGGTTC-----T	291
Db	250	AGTCGACAGACTGATGACACAGCCATCTACTACTGTGCGCAGAGATCATCTACTATGATT	309
Qy	292	ACGGTAGATTCTATGAGTACACTGCGGGGCCAAGGAGGACCGAGTCACC	336

Db 310 ACGGCTATGCTATGACTGAGGGGCGGCAAGGACGAGCTCACC 354  
RESULT 6  
LOCUS MUSIGHAXI 354 bp mRNA linear ROD 27-APR-1993  
DEFINITION Mouse Ig heavy-chain mRNA V region, partial cds from hybridoma H61-15VH.  
ACCESSION M36228  
VERSION M36228.1 GI:194839  
KEYWORDS V-region; immunoglobulin heavy chain; processed gene.  
SOURCE Mouse (strain Balb/c), cDNA to mRNA, from hybridoma H61-15.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 354)  
AUTHORS Kavalier,J., Caton,A.J., Staudt,L.M., Schwartz,D. and Gerhard,W.  
TITLE A set of closely related antibodies dominates the primary antibody response to the antigenic site CB of the A/PR/8/34 Influenza virus hemagglutinin  
JOURNAL J. Immunol. 145, 2312-2321 (1990)  
MEDLINE 90375932  
COMMENT Draft entry and computer-readable sequence for [J. Immunol. (1990) in press] kindly submitted by J.Kavalier, 06-JUL-1990.  
FEATURES  
source location/Qualifiers  
1..354  
/organism="Mus musculus"  
/strain="BALB/c"  
/sub\_species="domesticus"  
/db\_xref="taxon:10090"  
/map="chromosome 12"  
/cell\_line="H61-15"  
/tissue\_type="hybridoma"  
1..354  
/gene="Igh"  
/gene="Igh"  
<1..>354  
/codon\_start=1  
/product="Immunoglobulin heavy chain V-region"  
/protein\_id="AAA38067.1"  
/db\_xref="GI:194840"  
/translation="QVQLKESGPELVAPSSQSLTITCVSGESLTSTGVHWKPPCKG LEWLVGIVAGSGSTNSALMSRLSISKNSQVFLKMSLQTDITAMYYCARGHYYG YIAMDYGGQSTVVS"  
BASE COUNT 90 a 89 c 93 g 82 t  
ORIGIN  
Query Match 84.2%; Score 283; DB 10; Length 354;  
Best Local Similarity 91.7%; Pred. No. 1.7e-74;  
Matches 311; Conservative 0; Mismatches 25; Indels 3; Gaps 1;

Qy 298 GATTCTATGACTAGTGGGGCCAGGACGAGCTCACC 336  
Db 310 TATGCTATGACTAGTGGGGTCAAGGAACCTGATCACC 348  
RESULT 7  
LOCUS AX256252 357 bp DNA linear PAT 10-OCT-2001  
DEFINITION Sequence 22 from Patent WO0171005.  
ACCESSION AX256252  
VERSION AX256252.1 GI:16075153  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 357)  
AUTHORS Kufer,P., Riethmuller,G., Luterbuese,R., Borschert,K., Kischel,R., Mayer,W. and Hofmeister,R.  
TITLE Multifunctional polypeptides comprising a binding site to an epitope of the nk92d receptor complex  
Patent: WO 0171005-A 22 27-SEP-2001;  
JOURNAL Kufer, Peter (DE)  
FEATURES  
source location/Qualifiers  
1..357  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 85 a 87 c 104 g 81 t  
ORIGIN  
Query Match 83.3%; Score 279.8; DB 6; Length 357;  
Best Local Similarity 91.2%; Pred. No. 1.6e-73;  
Matches 309; Conservative 0; Mismatches 27; Indels 3; Gaps 1;

Qy 1 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGCGCTTCATCACTTGCAC 60  
Db 10 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGCGCTTCATCACTTGCAC 69  
Qy 61 GTCTCTGGGTTTTCATTAAACAGCTATGATGTATGCTGGTTCGCGACCTCCAGGAAG 120  
Db 70 GTCTCTGGGTTTTCATTAAACAGCTATGATGTATGCTGGTTCGCGACCTCCAGGAAG 129  
Qy 121 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAACACAAATTAATTGGCTCTC 180  
Db 130 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAACACAAATTAATTGGCTCTC 189  
Qy 181 ATGTCCAGACTGAACATCAACAGACAATTCACAGACCAATTTCTTAATAATGAAC 240  
Db 190 ATGTCCAGACTGAACATCAACAGACAATTCACAGACCAATTTCTTAATAATGAAT 249  
Qy 241 AGTCTCAAACTGATGACACAGCCATCTACTGTGCGAGAGAGGTT---CTACGGTA 297  
Db 250 AGTCTCAAACTGATGACACAGCCATCTACTGTGCGAGAGGGGGTACAGAGGGGG 309  
Qy 298 GATTCTATGACTAGTGGGGCCAGGACGAGCTCACC 336  
Db 310 GCTGTGTTGGTTACTGGGGCCAGGACGAGCTCACC 348  
RESULT 8  
LOCUS MUSIGHAXI 360 bp mRNA linear ROD 27-APR-1993  
DEFINITION Mouse Ig heavy-chain mRNA V region, partial cds. H220-22VH.  
ACCESSION M36217  
VERSION M36217.1 GI:194815  
KEYWORDS V-region; immunoglobulin heavy chain; processed gene.  
SOURCE Mouse (strain Balb/c), cDNA to mRNA, from hybridoma H220-22.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 360)  
AUTHORS Kavalier,J., Caton,A.J., Staudt,L.M., Schwartz,D. and Gerhard,W.



SOURCE immunoglobulin-kappa.  
Mus musculus (strain BALB/c, sub-species domesticus) female ascites  
CDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 345)  
AUTHORS Deng, J., Chua, M.M., Andrews, G.C. and Karush, F.  
TITLE Primary B-cell response to neuropeptide Y and bovine pancreatic  
polypeptide  
JOURNAL Mol. Immunol. 29 (7-8), 847-856 (1992)  
MEDLINE 92342150  
FEATURES  
source  
1..345  
location/Qualifiers  
/organism="Mus musculus"  
/strain="BALB/c"  
/sub\_species="domesticus"  
/db\_xref="taxon:10090"  
/map="chromosome 12"  
/sex="female"  
/cell\_line="DSP2.2"  
/cell\_type="hybridoma"  
/tissue\_type="ascites"  
1..345  
/gene="Igm"  
1..345  
/partial  
/gene="Igm"  
/codon\_start=1  
/product="mu-Immunoglobulin"  
/protein\_id="AA51159.1"  
/db\_xref="GI:197944"  
/translation="LESGPGIVAPQSLSITCTVSGSLSYGVHWVROPKGLLEML  
GLVWAGSTNNSALMSRLSKDSSKSYFLKMNLSIQTDITAMVYCA SPLITTYAM  
DYMGGTSVTVSS"  
4..278  
/gene="Igm"  
279..283  
/gene="Igm"  
284..291  
/gene="Igm"  
292..294  
/gene="Igm"  
295..345  
/partial  
/gene="Igm"  
BASE COUNT 86 a 90 c 84 g 85 t  
ORIGIN  
Query Match 82.0%; Score 275.4; DB 10; Length 345;  
Best Local Similarity 91.3%; Pred. No. 3.3e-72;  
Matches 304; Conservative 0; Mismatches 26; Indels 3; Gaps 1;

QY 304 ATGACTACTGGGCCAAGGACACAGCGTCACC 336  
|||||  
DB 304 ATGACTACTGGGTCACAGACCTCAGTCACC 336  
RESULT 11  
MMU60456  
LOCUS MMU60456 345 bp mRNA linear ROD 10-JUL-1996  
DEFINITION Mus musculus Ig anti-DNA heavy chain VDJ (VH Q52) mRNA, partial  
cds.  
ACCESSION U60456  
VERSION U60456.1 GI:1407739  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 345)  
AUTHORS Wioch, M.K., Alexander, A.L., Pippen, A.M.M., Lefkowitz, J.B.,  
Pisetsky, D.S. and Gillespie, G.S.  
TITLE Anti-DNA induced in Preautimmune NZB/W Mice by Bacterial DNA  
Immunization  
JOURNAL unpublished  
REFERENCE 2 (bases 1 to 345)  
AUTHORS Wioch, M.K., Alexander, A.L., Pippen, A.M.M., Lefkowitz, J.B.,  
Pisetsky, D.S. and Gillespie, G.S.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUN-1996) Research Service, Durham VA Med Center, 508  
Fulton Box 151G, Durham, NC 27705, USA  
FEATURES  
source  
1..345  
location/Qualifiers  
/organism="Mus musculus"  
/strain="(NZB x NZW)F1"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/clone="B cell hybridoma F4-1"  
1..345  
/gene="VH Q52"  
<1..>345  
/gene="VH Q52"  
/note="Description: VH gene; heavy chain variable region"  
/codon\_start=1  
/product="Ig heavy chain"  
/protein\_id="AAB03591.1"  
/db\_xref="GI:1407740"  
/translation="QVQLKESGPIVLPAPSOGLSITCTVSGSLSYGVHWVROPKGL  
LEMLGVWAGSTNNSALMSRLSKDSSKSYFLKMNLSIQTDITAMVYCAKOGVAM  
DYMGGTSVTVSS"  
BASE COUNT 89 a 86 c 91 g 79 t  
ORIGIN  
Query Match 81.7%; Score 274.4; DB 10; Length 345;  
Best Local Similarity 91.1%; Pred. No. 6.7e-72;  
Matches 306; Conservative 0; Mismatches 21; Indels 9; Gaps 1;

RESULT 12  
AF006578 342 bp mRNA linear ROD 12-JUL-1997  
LOCUS Mus musculus anti-DNA autoantibody heavy chain variable region  
DEFINITION mDNA, partial cds.  
ACCESSION AF006578  
VERSION AF006578.1 GI:2253323  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 342)  
AUTHORS Weller,S., Conde,C., Knapp,A.M., Levallois,H., Giffillan,S., Pasquali,J.L. and Martin,T.  
TITLE Autoantibodies in mice lacking terminal deoxynucleotidyl transferase: evidence for a role of N-addition in the polyreactivity and in the affinities of anti-DNA antibodies  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 342)  
AUTHORS Weller,S., Conde,C., Knapp,A.M., Levallois,H., Giffillan,S., Pasquali,J.L. and Martin,T.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1997) Laboratory of Immunopathology, Institute of Immunology, Central Hospital, 1 Place de l'hospital, Strasbourg 67000, France  
FEATURES  
source Location/Qualifiers  
1..342  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/chromosome="12"  
/cell\_line="spleen hybridoma cell line 64.3A11"  
/note="anti-DNA autoantibody; V-D-J region"  
/codon\_start=1  
/product="IgM heavy chain variable region"  
/protein\_id="AA062902.1"  
/translation="LDYEGPLVAPSGSLSTCTVSGFSLTYGVHWKPPGKGLM LGVIMAGSTNYNSALMSRLSISKDMSQVFLKMSLQSDPTAMXYCARGGIITYAM DHGRILVSRPG"  
BASE COUNT 84 a 87 c 89 g 82 t  
ORIGIN  
Query Match 81.6%; Score 274.2; DB 10; Length 342;  
Best Local Similarity 91.2%; Pred. No. 7.7e-72;  
Matches 291; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Db 181 ATGTCCAGACTGAGCATCAGCAAGCAACTCCAGAGCCAAAGTTTCTTAATAATGAAAC 240  
Qy 241 AGTCGCAACTGATGACACAGCCCATCTACTAGTGTGCAGAGAGGTTCTACGTTAGAT 300  
Db 241 AGTCGCAACTGATGACACAGCCCATCTACTAGTGTGCAGAGAGGTTCTACGTTAGAT 300  
Qy 301 TCTATGACTACTGGGGCC 319  
Db 301 GCTATGACTACTGGGGTCAAGAACTCACTCACC 336  
RESULT 13  
AR083821 351 bp DNA linear PAT 01-SEP-2000  
LOCUS AR083821  
DEFINITION Sequence 36 from patent US 5977316.  
ACCESSION AR083821  
VERSION AR083821.1 GI:10010592  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 351)  
AUTHORS Chatterjee,M., Foon,K.A. and Chatterjee,S.K.  
TITLE Monoclonal antibody 1A7 and related polypeptides  
JOURNAL Patent: US 5977316-A 36 02-NOV-1999;  
FEATURES  
source Location/Qualifiers  
1..351  
/organism="unknown"  
BASE COUNT 95 a 87 c 88 g 80 t 1 others  
ORIGIN  
Query Match 81.4%; Score 273.6; DB 6; Length 351;  
Best Local Similarity 88.4%; Pred. No. 1.2e-71;  
Matches 297; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 CTGCAGAGTCAGGACCTGGCTGTGGCGCCCTCAGAGAGCCTGTCATCATTGCACT 60  
Db 10 CTGAAGGAGTCAGGACCTGGCTGTGGCGCCCTCAGAGAGCCTGTCATCATTGCACT 69  
Qy 61 GTCTGTGGGTTTTCATTAAACAGCTATGCTGACACTGGTTCGCCAGCCTCCAGGAAG 120  
Db 70 GTCTGTGGGTTTTCATTAAACAGCTATGCTGACACTGGTTCGCCAGCCTCCAGGAAG 129  
Qy 121 GGTCTGAGTGGCTGGGAGTCAATTTGGGCTGTGGAGACAAATTTATTAATTCGCTTC 180  
Db 130 GGTCTGAGTGGCTGGGAGTCAATTTGGGCTGTGGAGACAAATTTATTAATTCGCTTC 189  
Qy 181 ATGTCCAGACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGTTCTTAAATGAAC 240  
Db 190 AAATCCAGACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGTTCTTAAATGAAC 249  
Qy 241 AGTCGCAACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGTTCTTAAATGAAC 300  
Db 250 AGTCGCAACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGTTCTTAAATGAAC 309  
Qy 301 TCTATGACTACTGGGGCCAGGAGGCAAGCTCACC 336  
Db 310 GCTATGACTACTGGGGTCAAGAACTCACTCACC 345  
RESULT 14  
MMSWS1H 351 bp mRNA linear ROD 23-FEB-1994  
LOCUS MMSWS1H  
DEFINITION M. musculus (A.SW) mRNA for ASMS1 antibody heavy chain variable region.  
ACCESSION X75099  
VERSION X75099.1 GI:414151  
KEYWORDS antibody; variable region; VH region.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 09:36:01 ; Search time 363.1 Seconds  
(without alignments)  
1588.772 Million cell updates/sec

Title: US-09-824-286-6

Perfect score: 336  
Sequence: 1 CTCGAGGAGTCAGGACCTG.....GCCAAGGAGCCAGGTACC 336

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_032802.\*  
1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	336	100.0	336	19	AA97441
2	302.4	90.0	759	24	AA597140
3	279.8	83.3	357	24	AA597129
4	275.2	81.9	360	22	AA78183
5	274	81.5	714	18	AA786310
6	274	81.5	1173	18	AA786312
7	273.6	81.4	351	20	AA31382
8	273.6	81.4	420	18	AA77852
9	271	80.7	1431	22	AA91017

10	271	80.7	1431	22	AA91018	Chimeric 4H6 anti-
11	270.2	80.4	357	21	AA38902	260P9 hybridoma VH
12	269.8	80.3	304	20	AA31378	MUSIGHAEI Mouse Ig
13	269.6	80.2	669	14	AA037462	VH NO2/12.4-VK NO1
14	269.6	80.2	708	14	AA037459	VH NO2/12.4-VK NO1
15	267.6	79.6	291	20	AA31383	MUSIGHAD mouse Ig
16	267.2	79.5	720	17	AA36463	18-2-3/TRY202' sin
17	267.2	79.5	720	17	AA31379	Single chain bind1
18	267.2	79.5	729	11	AA005715	18-2-3-/TRY59. A
19	267.2	79.5	729	14	AA051541	Coding sequence of
20	267.2	79.5	729	17	AA36464	18-2-3/TRY59 singl
21	267.2	79.5	729	17	AA313740	Single chain bind1
22	266.4	79.3	870	13	AA036982	Ox VH-hinge-VL Ins
23	265.6	79.0	724	14	AA051540	Coding sequence of
24	265.6	79.0	724	11	AA005714	18-2-3-/TRY202'
25	264.6	78.8	411	22	AA786854	Ganglioside GD2 sp
26	264	78.6	353	13	AA027175	H-chain variable r
27	263	78.3	411	18	AA790984	Nucleotide sequenc
28	260.8	77.6	796	21	AA50583	Murine anti-Lewis
29	260	77.4	334	17	AA34097	Mab 2B6 heavy cha1
30	259.8	77.3	359	21	AA27332	Murine consensus 1
31	259.6	77.3	292	20	AA31385	MUSIGHAF mouse IgH
32	259.4	77.2	369	18	AA743807	Anti-DNA antibody
33	259	77.1	333	19	AAV03489	Mouse anti-human I
34	258	76.8	292	20	AA31380	MUSAB M. musculus
35	258	76.8	293	20	AA31389	MUSIGHCVX mouse Ig
36	258	76.8	737	18	AA794606	ESGCR1-HindIII inse
37	258	76.8	737	18	AA766143	PUR.4124 insert en
38	258	76.8	3922	22	AA78408	Nucleotide sequenc
39	258	76.8	3922	22	AA78409	Nucleotide sequenc
40	256.4	76.3	639	16	AA076280	Truncated scfv ant
41	256.4	76.3	852	11	AA004467	Two linked VHLYS d
42	256.4	76.3	889	13	AA021096	Scfv sequences in
43	256.4	76.3	895	16	AA076272	Scfv anti-lysozyme
44	256.4	76.3	915	11	AA004466	Light chain variab
45	256.4	76.3	923	16	AA704171	Plasmid pSV1VHD1.3

#### ALIGNMENTS

RESULT 1	
AA797441	
ID	AA797441 standard; cDNA; 336 BP.
AC	AA797441:
XX	
DT	21-MAY-1998 (first entry)
XX	
DE	Monoclonal antibody CP.B8 heavy chain variable region cDNA.
XX	
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; CP.B8; immunological disease;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;
SS.	
XX	
OS	Mus musculus.
XX	
PN	WO9743416-A1.
PD	20-NOV-1997.
XX	
PF	09-MAY-1997; 97WO-US07870.
XX	
PR	10-MAY-1996; 96US-0017466.
XX	
PA	(BIOU ) BIOGEN INC.
XX	
PI	Benjamin CD, Burkly LC, Hession C, Whitty A;
XX	

DR WPI: 1998-008885/01.  
 DR P-PSDB: AAW31648.  
 XX  
 PT Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell anergy for  
 PT treatment of immunological diseases  
 PS  
 PS Claim 21; Page 82; 111pp; English.  
 XX  
 CC This cDNA sequence codes for the heavy chain variable region (see  
 CC AAW31648) of monoclonal antibody (Mab) C9.88, which is produced  
 CC by a hybridoma deposited as ATCC 12107, and which is specific for  
 CC the gamma constant (gc) chain (see AAW31646) of human cytokine  
 CC receptors. A Mab having complementarity determining regions  
 CC encoded by the VH or VL (see AAT07440) cDNA is claimed. The  
 CC invention provides compositions and methods for inhibiting cytokine  
 CC signalling using gc chain blocking agents for the treatment of  
 CC immunological diseases such as myasthenia gravis, rheumatoid  
 CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
 CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
 CC allergy, asthma, parasitic infection, graft vs. host disease or  
 CC psoriasis. Nucleic acids encoding the gc-blocking agents can be  
 CC used in gene therapy for the same purposes. A preferred gc  
 CC blocking agent in Mab CP.88 or its Fab fragment.  
 XX  
 SQ Sequence 336 BP; 85 A; 84 C; 88 G; 79 T; 0 other;

Query Match 100.0%; Score 336; DB 19; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-93;  
 Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCAGAGTGTAGACCTGGCTGGTGGCCCTCACAGAGCTGTCCATCTTGCACCT 60  
 |||  
 Db 1 ctgcagagagtcagacccgctggtgctgcctcaccagagcgttcacatctgcact 60  
 QY 61 GTCTCGGCTTTTCATTAAACAGCTATGCTGCTGGTTCGCCACCTCCAGGAAG 120  
 |||  
 Db 61 gtctcggcttttcattaaacagctatgctgctgggttcgccaccctccaggaaag 120  
 QY 121 GGTCTGAGTGGCTGGAGTCAATTTGGGCTGTGGAGACACAAATTAATTCGGCTTC 180  
 |||  
 Db 121 ggtctgagtggtcggagagtcatttggctgtggtgaagcacaaattatcgtcttc 180  
 QY 181 ATGTCGACCTGACATCAACAGACATTCACAAAGCCAAATTTTCTTAAATGAC 240  
 |||  
 Db 181 atgtccgagctgacaatcaacagagacaattccaagagccaatttctttaaattgac 240  
 QY 241 AGTCTGCAACTGATGACAGACCATCTACTACTGTGTCAGAGAGGTTCTACGGTAGAT 300  
 |||  
 Db 241 agtctgcaactgatacagacagcatctactactcgtgccaagaggttctacggtgat 300  
 QY 301 TCTATGACTACTGTGGGCCAAGGACACAGGTCAAC 336  
 |||  
 Db 301 tctatgactactgtgggccaagagagccaggtcac 336

RESULT 2  
 AAS97140  
 ID AAS97140 standard; DNA: 759 BP.

XX AAS97140;  
 XX 26-FEB-2002 (first entry)  
 XX  
 DE P5-3 single chain Fv DNA.  
 XX  
 KW Human: NKG2D receptor complex; cancer; infectious disease; tumour;  
 KW autoimmune disease; head; neck; stomach; oesophagus; colon; liver; ovary;  
 KW intrahepatic bile duct; pancreas; lung; larynx; breast; uterus; cervix;  
 KW prostate; kidney; testis; thyroid; bladder; brain; melanoma; myeloma; Fv;  
 KW sarcoma; leukaemia; lymphoma; virus; bacterium; fungus; protozoan; ds;  
 KW helminth; cytosstatic; antimicrobial; immunomodulatory; P4-2; P4-3; P4-14;

KW P4-15; P5-2; P5-3; P5-9; P5-10; P5-11; P5-23; 3B10xP4-3; 3B10xP4-14;  
 KW P53 tetramerisation domain; 3B10xP5-2; 3B10xP5-23.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200171005-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 26-MAR-2001; 2001WO-EP03414.  
 PF  
 XX 24-MAR-2000; 2000EP-0106467.  
 PR  
 XX (KUFE/) KUFER P.  
 PA  
 PI Kufner P, Rietmueller G, Lutterbuese R, Borschert K, Kischel R;  
 PI Mayer M, Hofmeister R;  
 PT  
 XX WPI: 2002-055119/07.  
 DR P-PSDB: AAU72866.

Multifunctional polypeptides comprising binding sites that specifically  
 recognise extracellular groups of the NKG2D receptor complex and  
 PT domains which function as receptors or ligands, useful for treating  
 cancers and infectious diseases -

Example 7; Fig 16; 114pp; English.

XX The invention relates to a multifunctional polypeptide comprising a  
 CC domain with a binding site that specifically recognises an extracellular  
 CC group of the NKG2D receptor complex and a second domain which functions  
 CC as a receptor or ligand. The polypeptide and its associated  
 CC polynucleotide are used for the preparation of a pharmaceutical  
 CC composition for the treatment of cancer, infections and/or autoimmune  
 CC conditions. The cancer may be a tumour of the head and neck, stomach,  
 CC oesophagus, colon, liver, intrahepatic bile ducts, pancreas, lung,  
 CC larynx, breast, ovary, uterus, cervix, prostate, kidney, testis, thyroid,  
 CC bladder or brain, or a melanoma, myeloma, sarcoma, leukaemia or lymphoma.  
 CC The infectious diseases can be caused by viruses, bacteria, fungi,  
 CC protozoa or helminths. The autoimmune diseases include multiple  
 CC sclerosis, Grave's disease, ankylosing spondylitis, acute anterior  
 CC uveitis, Goodpasture's syndrome, myasthenia gravis, insulin-dependent  
 CC diabetes mellitus, rheumatoid arthritis, pemphigus vulgaris and  
 CC autoimmune hepatitis. The sequences represent DNA encoding the  
 CC polypeptides of the invention.  
 XX  
 SQ Sequence 759 BP; 188 A; 186 C; 201 G; 184 T; 0 other;

Query Match 90.0%; Score 302.4; DB 24; Length 759;  
 Best Local Similarity 93.8%; Pred. No. 4.9e-83;  
 Matches 315; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CTGCAGAGTGTAGACCTGGCTGGTGGCCCTCACAGAGCTGTCCATCTTGCACCT 60  
 |||  
 Db 13 ctgcagagagtcagacccgctggtgctgcctcaccagagcgttcacatctgcact 72  
 QY 61 GTCTCGGCTTTTCATTAAACAGCTATGCTGCTGCTGGTTCGCCACCTCCAGGAAG 120  
 |||  
 Db 73 gtctcggcttttcattaaacagctatgltacactgtgtccagcctccaggaaag 132  
 QY 121 GGTCTGAGTGGCTGGAGTCAATTTGGGCTGTGGAGACACAAATTAATTCGGCTTC 180  
 |||  
 Db 133 ggtctgagtggtcggagagtaataatggtggtggtggtggtggtggtggtggtggt 192  
 QY 181 ATGTCGACCTGACATCAACAGACAGCAATTCACAAAGCCAAATTTTCTTAAATGAC 240  
 |||  
 Db 193 atgtccgagctgagacatcagaagacacaccccaagagccaagtttctttaaattgac 252  
 QY 241 AGTCTGCAACTGATGACACAGACCATCTACTACTGTGTCAGAGAGGTTCTACGGTAGAT 300  
 |||  
 Db 253 agtctgcaactgatacagacagcatctactactcgtgccaagagatcggtactacgtggt 312

OY 301 TCATGACACTACGGGGCCAAAGACACGGTCACC 336  
 Db 313 gctatggactactgtgggccaaggaccacgctcacc 348

## RESULT 3

AAS97129  
 ID AAS97129 standard; cDNA: 357 BP.

XX AAS97129;

AC 26-FEB-2002 (first entry)

DE Anti-NG2D hybridoma 6H7E7 variable heavy chain DNA.

XX Human: NG2D receptor complex; cancer; infectious disease; tumour;  
 KW autoimmune disease; head; neck; stomach; oesophagus; colon; liver; ovary;  
 KW intrahepatic bile duct; pancreas; lung; larynx; breast; uterus; cervix;  
 KW prostate; kidney; testis; thyroid; bladder; brain; melanoma; myeloma; ss;  
 KW sarcoma; leukaemia; lymphoma; virus; bacterium; fungus; protozoan; DAPI10;  
 KM helminth; cytostatic; antimicrobial; immunomodulatory; 11B2D10; 6H7E7;  
 KM 867C10; 6E5A7; PCR primer.

OS Homo sapiens.

XX MO200171005-A2.

XX 27-SEP-2001.

XX 26-MAR-2001; 2001WO-EP03414.

XX 24-MAR-2000; 2000EP-0106467.

XX (KUFE/) KUFER P.

XX Kufer P, Riettmueller G, Lutterbuese R, Borschert K, Kischel R;  
 PI Mayer M, Hofmeister R;

XX WPI: 2002-055119/07.

XX P-PSDB; AAU72837.

PT Multifunctional polypeptides comprising binding sites that specifically  
 PT recognise extracellular groups of the NG2D receptor complex and  
 PT domains which function as receptors or ligands, useful for treating  
 PT cancers and infectious diseases -

XX Example 3; Fig 16; 114pp; English.

XX The invention relates to a multifunctional polypeptide comprising a  
 CC domain with a binding site that specifically recognises an extracellular  
 CC group of the NG2D receptor complex and a second domain which functions  
 CC as a receptor or ligand. The polypeptide and its associated  
 CC polynucleotide are used for the preparation of a pharmaceutical  
 CC composition for the treatment of cancer, infections and/or autoimmune  
 CC conditions. The cancer may be a tumour of the head and neck, stomach,  
 CC oesophagus, colon, liver, intrahepatic bile ducts, pancreas, lung,  
 CC larynx, breast, ovary, cervix, prostate, kidney, testis, thyroid,  
 CC bladder or brain, or a melanoma, myeloma, sarcoma, leukaemia or lymphoma.  
 CC The infectious diseases can be caused by viruses, bacteria, fungi,  
 CC protozoa or helminths. The autoimmune diseases include multiple  
 CC sclerosis, Grave's disease, ankylosing spondylitis, acute anterior  
 CC uveitis, Goodpasture's syndrome, myasthenia gravis, insulin-dependent  
 CC diabetes mellitus, rheumatoid arthritis, pemphigus vulgaris and  
 CC autoimmune hepatitis. The sequences represent the NG2D receptor DNA, DNA  
 CC encoding the polypeptides of the invention and PCR primers used to  
 CC amplify the DNA sequences.

XX Sequence 357 BP; 85 A; 87 C; 104 G; 81 T; 0 other;

## Query Match

Best Local Similarity 83.3%; Score 279.8; DB 24; Length 357;  
 Matches 309; Conservative 0; Mismatches 27; Indels 3; Gaps 1;

OY 1 CTGCAGAGTACAGACCTGGCCCTGGCCCTTCACAGACCTGTCATCATCTTGCACT 60  
 Db 10 ctgcagaggtacagaccctggcctggtggccctcacagacctgtccatcacttgcaact 69  
 OY 61 GTCTGTGGCTTTTCATTAAACGACTAGTGTGTACCTGGTTCCGACCTCCAGGAAG 120  
 Db 70 gtctgtgggttttcattaaacgactatgtgtacactgattgcgcagctccaggaag 129  
 OY 121 GGTCTGAGTGGCTGGGAGTCATTTGGGCTGGTGAACACAAATTAATTCGGCTTC 180  
 Db 130 ggtctgagtggtgggagtcatttgggctgggtgaacacaaatataattcggcttc 189  
 OY 181 ATGTCACAGCTGACATCAACAGACAAATTCACAGACCCAAATTTCTTAAAAATGAAC 240  
 Db 190 atgtcacagctgagatcaagaagacaaccccaagacaaagtttcttaaaatgat 249  
 OY 241 AGTCTCAAACTGATGACACAGCCATCTACTAGTGTGCCAGAGAGGTTT--CTACGGTA 297  
 Db 250 agtctcaaatgtatgacacagccatgtactactgtgccagaggggttacgaaggcg 309  
 OY 298 GATTCTATGACTACTGGGGCCCAAGGACACACGGTCACC 336  
 Db 310 gctgtgtgtgtactgtgggccaaggaccacgctcacc 348

## RESULT 4

AAH78183  
 ID AAH78183 standard; DNA: 360 BP.

XX AAH78183;

XX 26-NOV-2001 (first entry)

DE Nucleotide sequence of heavy chain variable region of antibody 2C4.

XX Monoclonal antibody 2C4; sialoadhesin factor-2; SAF-2; allergic rhinitis;  
 KW allergy; asthma; anemia; eczema; lymphoma; systemic mastocytosis;  
 KW leukemia; eosinophil; ds.

XX Mus sp.

XX MO200166126-A1.

XX 13-SEP-2001.

XX 05-MAR-2001; 2001WO-US07193.

XX 07-MAR-2000; 2000US-0187595.

XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PA (UTJO ) UNIV JOHNS HOPKINS.

XX Abrahamson JA, Bochner B, Erickson-Miller CL, Kikly KK;  
 PI Schleimer R;

XX WPI: 2001-570749/64.

XX P-PSDB; AAG65986.

PT Novel monoclonal antibody specific for human sialoadhesin factor-2 for  
 PT diagnosis, prevention, treatment of allergy, asthma, eczema or diseases  
 PT such as lymphoma, leukemia or systemic mastocytosis, in a mammal -  
 XX Disclosure; Fig 1; 35pp; English.

XX The present sequence encodes the heavy chain variable region of murine  
 CC monoclonal antibody 2C4. This antibody binds to human sialoadhesin  
 CC factor-2 (SAF-2). The antibody is useful for treating or preventing  
 CC allergic rhinitis, allergies, asthma, anemia, eczema or diseases such  
 CC as lymphoma, leukemia or systemic mastocytosis in a mammal. It is also  
 CC useful for detecting the presence of a cell, especially eosinophil in  
 CC a sample, by detecting binding of the antibody to SAF-2. The antibody

CC can be coupled to toxins, antiproliferative drugs or radionuclides to  
 CC kill cells in areas of excessive SAF-2 expression.  
 XX  
 SQ Sequence 360 BP; 94 A; 92 C; 88 G; 86 T; 0 other;

Query Match 81.9%; Score 275.2; DB 22; Length 360;  
 Best Local Similarity 90.1%; Pred. No. 8,6e-75;  
 Matches 308; Conservative 0; Mismatches 28; Indels 6; Gaps 1;

OY 1 CTGACAGACTGAGACCTGGCTGGTGGCCCTCAGACAGCCTGTGCATCTTGCAT 60  
 || |||||  
 Db 10 ctaagagagctcagagcctgctgtgagccctcacagagcctgtccatctgcat 69  
 OY 61 GTCCTGGGTTTTCATTAACACGCTATGCTACACTGGGTTGCCAGCTCCAGAAAG 120  
 |||||  
 Db 70 gctctcgggttctcattcaaccatctatgctcactggtctgcagagcctccaggaag 129  
 OY 121 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGGAAAGCAAAATTAATTCGGCTTC 180  
 |||||  
 Db 130 ggtctggagtgctggagtaataatggtcgtggaagacataataatcgcctc 189  
 OY 181 ATGTCAGACTGACATCAACAGAGCAATTCACAGACCAATTTCTTAAATAATGAC 240  
 |||||  
 Db 190 atgtccagactgagcactcagcaagaacactccaaagagccttcttcttaaaataaac 249  
 OY 241 ACTTCGCAACACTGACACACCATCTACTGTGCCAGAGAGGTTCTA-----CG 294  
 |||||  
 Db 250 agctcgcaaacactgatacagcagccctgtactctgtccagagagctagccctat 309  
 OY 295 GTAGATTCTATGACTACTGGGCGCCAGGACACGACGTCAC 336  
 |||||  
 Db 310 tactattctatgatactcgtgggtcaagaacctcagtcacc 351

RESULT 5  
 AAT86310  
 ID AAT86310 standard; DNA; 714 BP.

AC AAT86310;

DT 06-APR-1998 (first entry)

DE Single chain anti-disialoganglioside GD2 antibody 3G6-scfv.

KM Antibody construct; disialoganglioside; GD2; single chain Fv fragment;

KM scfv; tumour; neuroblastoma; osteosarcoma; soft tissue sarcoma;

KM tissue imaging; target delivery; toxin; streptavidin;

KM pro-drug converting enzyme; GD2-targeted lymphocyte; ss.

OS Synthetic.

FN WO9734634-A1.

PD 25-SEP-1997.

PE 20-MAR-1997; 97WO-US04427.

PR 20-MAR-1996; 96US-0013703.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PI Cheung NV, Guo H, Larson SM, Rivlin K, Sadelain M;

DR WPI; 1997-479996/44.

PT Recombinant single chain anti-disialoganglioside GD2 antibody -

PT useful to detect tumour cells expressing GD2 and to target

PS therapeutic agents, e.g. toxins, to such cells

PS Disclosure; Page 11-12; 31pp; English.

CC The present sequence encodes a recombinant single chain peptide,

CC 3G6-scfv. The peptide is an antibody construct comprising the variable  
 CC regions of the heavy and light chains of an antibody against  
 CC disialoganglioside (GD2) as a single chain Fv fragment (scfv). GD2 occurs  
 CC in many tumours types including neuroblastoma, osteosarcomas and other  
 CC soft tissue sarcomas, medulloblastomas, high grade astrocytomas,  
 CC melanomas and small cell lung cancer. The peptide can be detectably  
 CC labelled, preferably with 99m-Tc, for tissue imaging of cells expressing  
 CC GD2. It can also be used to target delivery of a therapeutic or  
 CC pre-therapeutic agent, such as a toxin, streptavidin or a pro-drug  
 CC converting enzyme, to cells expressing GD2. The peptide may further  
 CC comprise CD8 to facilitate the formation of GD2-targeted lymphocytes.  
 CC T cells containing the peptide can also be used to target GD2-producing  
 CC tumour cells.

SQ Sequence 714 BP; 174 A; 168 C; 205 G; 167 T; 0 other;

Query Match 81.5%; Score 274; DB 18; Length 714;  
 Best Local Similarity 90.2%; Pred. No. 2,6e-74;  
 Matches 305; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

OY 2 TGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGACAGCCTGTGCATCTTGCATCG 61  
 || |||||  
 Db 368 tgaagagatcagagcctgctgtgagccctcacagagcctgtccatctgcatc 427  
 OY 62 TCTCTGGGTTTTCATTAACACGCTATGCTACACTGGGTTGCCAGCTCCAGAAAG 121  
 |||||  
 Db 428 tctctggttctcattcaaccatctatgctcactggtctgcagagcctccaggaag 487  
 OY 122 GTCCTGAGTGGCTGGAGTCAATTTGGCTGGTGGAAAGCAAAATTAATTCGGCTTC 181  
 |||||  
 Db 488 gctctggagtgctggagtaataatggtcgtggaagacataataatcgcctta 547  
 OY 182 TGTCCAGACTGACATCAACAGAGCAATTCACAGACCAATTTCTTAAATAATGACA 241  
 |||||  
 Db 548 tgtccagactgagcactcagcaagaacactccaaagagccttcttcttaaaataaac 607  
 OY 242 GTCCTGCAACACTGACACACCATCTACTGTGCCAGAGAGGTTCTA-----GAG 298  
 |||||  
 Db 608 gctctgcaaacactgatacagcagccctgtactctgtccagagagctagccctat 667  
 OY 299 ATTCTATGACTACTGGGCGCCAGGACACGACGTCAC 336  
 || |||||  
 Db 668 atgcttggactcactggtgggtcaagaacctcagtcacc 705

RESULT 6  
 AAT86312  
 ID AAT86312 standard; DNA; 1173 BP.

AC AAT86312;

DT 06-APR-1998 (first entry)

DE Single chain anti-disialoganglioside GD2 antibody 3G6-scfv-streptavidin.

KM Antibody construct; disialoganglioside; GD2; single chain Fv fragment;

KM scfv; tumour; neuroblastoma; osteosarcoma; soft tissue sarcoma;

KM tissue imaging; target delivery; toxin; streptavidin;

KM pro-drug converting enzyme; GD2-targeted lymphocyte; ss.

OS Synthetic.

FN WO9734634-A1.

PD 25-SEP-1997.

PE 20-MAR-1997; 97WO-US04427.

PR 20-MAR-1996; 96US-0013703.

PA (SLOK) SLOAN KETTERING INST CANCER RES.

PI Cheung NV, Guo H, Larson SM, Rivlin K, Sadelain M;

DR WPI; 1997-479996/44.

PT Recombinant single chain anti-disialoganglioside GD2 antibody -

PT useful to detect tumour cells expressing GD2 and to target

PS therapeutic agents, e.g. toxins, to such cells

PS Disclosure; Page 11-12; 31pp; English.

CC The present sequence encodes a recombinant single chain peptide,

FT /note="encodes streptavidin"  
XX  
PN W09734634-A1.  
XX  
PD 25-SEP-1997.  
XX  
PF 20-MAR-1997; 97MO-US04427.  
XX  
PR 20-MAR-1996; 96US-0013703.  
XX  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
PI Cheung NY, Guo H, Larson SM, Rivlin K, Sadelain M;  
XX WPI; 1997-479996/44.  
DR  
XX  
PT Recombinant single chain anti-distalogsanglioside GD2 antibody -  
PT useful to detect tumour cells expressing GD2 and to target  
PT therapeutic agents, e.g. toxins, to such cells  
XX  
XX  
PS Disclosure; Page 13; 31pp; English.  
XX  
CC The present sequence encodes a recombinant single chain peptide,  
CC 366-scFv-streptavidin. The peptide is an antibody construct comprising  
CC the variable regions of the heavy and light chains of an antibody against  
CC distalogsanglioside (GD2) as a single chain Fv fragment (scFv). GD2 occurs  
CC in many tumours types including neuroblastoma, osteosarcomas and other  
CC soft tissue sarcomas, medulloblastomas, high grade astrocytomas,  
CC melanomas and small cell lung cancer. The peptide can be detectably  
CC labelled, preferably with 99m-Tc, for tissue imaging of cells expressing  
CC GD2. It can also be used to target delivery of a therapeutic or  
CC pre-therapeutic agent, such as a toxin, streptavidin (e.g. present  
CC sequence) or a pro-drug converting enzyme, to cells expressing GD2. The  
CC peptide may further comprise CD8 to facilitate the formation of  
CC GD2-targeted lymphocytes. T cells containing the peptide can also be used  
CC to target GD2-producing tumour cells.  
XX  
SQ Sequence 1173 BP; 270 A; 334 C; 340 G; 229 T; 0 other;

Query Match 81.5%; Score 274; DB 18; Length 1173;  
Best Local Similarity 90.2%; Pred. No. 3.2e-74;  
Matches 305; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

QY 2 TGCAGAGTCAGACCTGGCGCTGGCGCCCTCACAGAGCCCTGCATCTGCACCTG 61  
DB 368 tgaagaggtcaagagccctgctgctgctccacagagccctgcatcctgcacg 427  
QY 62 TCTCTGGGTTTTCATTAAACAGCTATGCTGTACACTGGGTTCCGACCTCCAGGAAAG 121  
DB 428 tctctgggtttcatttaaccattatgltgacactggttcgcacgctccagaaag 487  
QY 122 GTCTGGAGTGGCTGGGAGTCAATTGGGCTGGTGAACACAAATTATATTCGGCTCA 181  
DB 488 gtctggagtgctgggagtaataatgctggtgaaagacaaataataatcgcctcta 547  
QY 182 TGTCCAACTGAACTCAACAGAGCAATTCAGAGCCAAATTTCTTAAATGAACA 241  
DB 548 tgtccaactgagatcaagcaagcaaccccaagagccaaagtttcttaaaatgaaca 607  
QY 242 GTCTGCAAACTGATGACACAGCCATCTACTACTGTGCCA--GAGAGGGTTCTACGGTAG 298  
DB 608 gtctgcaaaactgatacagacagccatgtaactactgtgccagcgggggttaactacgact 667  
QY 299 ATTCTATGAGTACTGGGGCCAAAGGACCAAGCTCACC 336  
DB 668 atgcttgactactggtggtcaagaaagaaactcaagtcacc 705

RESULT 7  
AAZ31382  
ID AAZ31382 standard; DNA: 351 BP.  
XX

AC AAZ31382;  
XX  
DT 07-FEB-2000 (first entry)  
XX  
DE MMASWSIH M. musculus A.SW mRNA for ASWSI antibody heavy chain.  
XX  
KW Monoclonal antibody; Mab; 1A7; GD2; immune response; melanoma;  
KW neuroblastoma; glioma; soft tissue carcinoma; small cell carcinoma;  
KW tumor-associated antigen; ss.  
XX  
OS Mus sp.  
XX  
PN US5977316-A.  
XX  
PD 02-NOV-1999.  
XX  
PF 16-JAN-1996; 96US-0591196.  
XX  
PR 17-JAN-1995; 95US-0372676.  
XX  
PA (KENT ) UNIV KENTUCKY.  
XX  
PI Foon KA, Chatterjee SK, Chatterjee M;  
XX  
DR WPI; 1999-619711/53.  
XX  
PT Monoclonal antibody 1A7 which elicits an anti-GD2 immunological  
PT response, useful for the development of products for the detection and  
PT treatment of cancers -  
XX  
XX  
PS Disclosure; Fig 13B; 74pp; English.  
XX  
CC The invention provides a monoclonal antibody (Mab) designated 1A7, which  
CC elicits an anti-GD2 (tumor-associated antigen) immunological response in  
CC humans. Mab 1A7 has defined light and heavy chain variable region  
CC sequences. The Mab 1A7 and polypeptides can be used for eliciting an  
CC anti-GD2 immune response. The polypeptides can also be used for  
CC detecting or purifying anti-GD2 antibody. The products can be used for  
CC treating GD2-associated diseases, e.g. melanoma, neuroblastoma, glioma,  
CC soft tissue carcinoma, and small cell carcinoma. They can be used for  
CC palliating the disease or for reducing the risk of recurrence. Sequences  
CC AAZ31373-90 represent sequences that closely matched Mab 1A7 heavy chain  
CC variable region encoding sequence.  
XX  
SQ Sequence 351 BP; 95 A; 87 C; 88 G; 80 T; 1 other;

Query Match 81.4%; Score 273.6; DB 20; Length 351;  
Best Local Similarity 88.4%; Pred. No. 2.6e-74;  
Matches 297; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 CTGCAGAGTCAGACCTGGCGCTGGCGCCCTCACAGAGCCCTGCATCTGCACCTG 60  
DB 10 ctgaagaggtcaagagccctgctgctgctccacagagccctgcatcactgcact 69  
QY 61 GTCTCTGGGTTTTCATTAAACAGCTATGCTGTACACTGGGTTCCGACCTCCAGGAAAG 120  
DB 70 gtctctgggtttcatttaaccattatgltgacactggttcgcacgctccagaaag 129  
QY 121 GGTCTGAGTGGCTGGGAGTCAATTGGGCTGGTGAACACAAATTATATTCGGCTTC 180  
DB 130 ggtctgagtgctgggagatgatatggtggtgaaacacagactaataatcagctctc 189  
QY 181 ATGTCCAGACTGAACTCAACAGAGCAATTCAGAGCCAAATTTCTTAAATGAAC 240  
DB 190 aaatccagactgagatcaagcaagcaaccccaagagccaaagtttcttaaaatgaac 249  
QY 241 AGCTGCAAACTGATGACACAGCCATCTACTACTGTGCCAAGAGGGTTCTACGGTAGAT 300  
DB 250 agctgcaaaactgatacagacagccatgtaactactgtgccagagagatggttaactacgactat 309  
QY 301 TCTATGACTACTGGGGCCAAAGGACCAAGCTCACC 336  
XXXXXXXXXXXXXXXXXXXXX

Db	310	gctatggactctcgtgggtcaagaacctagtcacc	345
		RESULT	8
		AAT77852	
ID	AAT77852	standard; cDNA;	420 BP.
XX			
AC	AAT77852;		
XX			
DT	03-NOV-1997	(first entry)	
XX			
DE	Murine anti-human class II monoclonal antibody 44H104 HL chain cDNA.		
XX			
KW	Antibody; heavy chain; variable region; hybridoma cell line 44H104;		
RN	immune response; enhance; stimulate; vaccine; immunodiagnosis;		
KM	antigen delivery; ss.		
XX			
OS	Mus musculus.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..420	
FT		/tag= a	
FT		/note= "Encodes 44H104 heavy chain variable region,	
FT		including secretion signal; termination	
FT		codon not given"	
PN	WO9640941-A1.		
XX			
PD	19-DEC-1996.		
XX			
PF	07-JUN-1996;	96WO-CAD00400.	
PR	07-JUN-1995;	95US-0483576.	
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
PI	Anand NN, Barber BH, Caterini JE, Cates GC, Klein MH;		
XX			
DR	WPJ; 1997-077271/07.		
DR	P-PSDB; AAM22538.		
XX			
PT	Recombinant conjugate antibody mol., modified for delivering an		
PT	antigen - elicits enhanced immune response without the use of		
PT	adjuvant to generate antibodies which are useful in vaccines or		
PT	immuno:diagnosis		
XX			
FS	Example 1; Fig 1B; 64pp; English.		
XX			
CC	Novel recombinant conjugate antibody molecules comprise a monoclonal		
CC	antibody specific for a surface structure of antigen presenting		
CC	cells (APC), genetically modified to contain at least one antigen		
CC	exclusively at one or more preselected sites. The conjugate is capabl		
CC	of delivering the antigen to APC and eliciting an immune response to		
CC	the antigen. The new conjugates are useful as vaccines and are able		
CC	to elicit an enhanced immune response without the use of an adjuvant.		
CC	In a specific example, a conjugate was constructed using the murine		
CC	anti-human class II monoclonal antibody secreted by hybridoma		
CC	44H104. The peptide CLR36 was chosen as antigen; it consists of		
CC	a tandemly linked T and B cell epitope derived from HIV MN strain.		
CC	The present sequence encodes the heavy chain variable region which		
CC	was PCR amplified from 44H104 and used in the preparation of a		
CC	conjugate with antigen CLR36.		
XX			
SQ	Sequence 420 BP; 95 A; 111 C; 108 G; 106 T; 0 other;		
	Query Match	81.4%; Score 273.6; DB 18; Length 420;	
	Best Local Similarity	89.8%; Pred. No. 2.8e-74;	
	Matches 307; Conservative	0; Mismatches 29; Indels 6; Gaps	
OY	1 CTGCGAGGTGACGACCGCTGGTGGGCCCTCACAGAGCCGTGCATCATTCGACT	60	
DB	70 ctgaagggttcaagaccccgctcgttggtgcctccacagagcctgtccatcaccttgact	1299	

OY		61	GTCCTGGGTTTTCATTAAACAGACTAATGGTGTAACACTGGGTGCCACAGCTCCAGGAAG	120
Dd		130	gctcttcgggtttccttaacagccatagtgtagactggtgtcgcagcctccaggaaag	189
OY		121	GGTCGAGAGTGGCTGGGAGTCATTGGCGCTGGTGGACACAATAATTCAATTCGGCTGC	180
Dd		190	ggctcggaagtgcgtggagtaatacatggcctggctggaaacaataattcaattcggctcc	249
OY		181	ATGTCACAGACTGACATCACAACAGAGACAATTCCAAGAGCCAAATTTCTTAAAAATGAAC	240
Dd		250	atgtcacagactgagcatcagcaagaacaactcaagagccaagtcttcttaaaaaatgagc	309
OY		241	ACTGTGCAAACTGATGACACACCACCTCTACTGTGGCAGAGAGGGTTCTACGGTAG--	298
Dd		310	agtcgtcaaacctgagacaacagcactgtactactgtgcagagacctatggtactacgtc	369
OY		299	---ATTCTATGACTACTGGGCGCCAGAGGACACCGCTCAC	336
Dd		370	cactatgtatgactactctgggtccaagaaacctcaagtcacc	411
Db				
RESULT 9				
ID	AAC91017			
ID	AAC91017 standard; DNA; 1431 BP.			
XX	AAC91017;			
AC				
XX				
DT	15-MAR-2001 (first entry)			
DE				
XX	Chimeric 4H6 anti-DR4 antibody heavy chain DNA.			
KM	Anti-Death receptor 4; DR4; antibody; apoptosis; cancer; arthritis;			
KW	autoImmune; ds.			
OS	Homo sapiens.			
OS	Synthetic.			
XX				
PN	WO200073349-A1.			
XX				
PD	07-DEC-2000.			
XX				
Pf	25-MAY-2000; 2000WO-US14599.			
XX				
PR	28-MAY-1999; 99US-0322875.			
XX	(GETH ) GENENTECH INC.			
PA				
PI	Ashkenazi AJ, Chuntharapai A, Dodge KH, Kim KJ;			
XX				
DR	WPI; 2001-041145/05.			
PT				
XX	Novel anti-death receptor 4 antibodies useful for treating cancer and			
PS	immune related disorders such as Rheumatoid arthritis, Sjogren's			
XX	sndrome, Grave's disease and diabetes mellitus -			
XX	Disclosure; Fig 18; 126pp; English.			
CC	The present invention relates to an anti-death receptor 4 (DR4)			
CC	antibody. The antibodies of the invention are useful for inducing			
CC	apoptosis in mammalian cancer cells such as colon cancer cells and			
CC	for treating an immune-related disease in a mammal such as arthritis			
CC	and autoimmune disease.			
XX				
SO	Sequence 1431 BP; 350 A; 438 C; 367 G; 273 T; 3 other:			
Query Match 80.7%; Score 271; DB 22; Length 1431;				
Best Local Similarity 96.5%; Pred.No.2.9e-73;				
Matches 277; Conservative 0; Mismatches 10; Indels 0; Gaps 0;				
OY	1 CTGCAAGAGTACGAGACCTGGCTGGTGGCGCCCTCACAGAGCGCTGTCCATCATCTTGCACT	60		

Db 67 ctgaagagtcacgacctggtcgtgctgcacctcacagagccctgcatactgacct 126  
QY 61 GTCCTGGGTTTTCATTAAACAGCTATGTTACACTGGGTTGGCCAGCTCCAGGAAG 120  
Db 127 gctccgggtcttcaataaccactatgctgtaacatcggtctgcagcccaagaa 186  
QY 121 GGTCTGGAGTGGCTGGGAGTCAATTTGGGCTGTGTGAAGCACAATTTAAATTCGGCTTC 180  
Db 187 gctctgagtgctggaataatagtgctgcttggaagacaataatatactgctctc 246  
QY 181 ATGTCCAGACTGAACATCAACAGACAGCAATTCACAAGCCAAATTTCTTAAAAATGAAAC 240  
Db 247 atgtccagactgagcactcaagaacactcaagaagccaagttcttcttaaaatgaa 306  
QY 241 AGTCTGAAACTGATGACAGCCATCTACTGTGCTCCAGAGAGG 287  
Db 307 agtctgcaactgatatgacagcaatgtaactatgctgctgcaagaaggg 353

RESULT 10  
AAC91018/c  
ID AAC91018 standard; DNA: 1431 BP.

XX AAC91018;  
AC  
XX 15-MAR-2001 (first entry)  
XX DT  
XX DE Chimeric 4H6 anti-DR4 antibody heavy chain complementary DNA.  
XX KW Anti-Death receptor 4; DR4; antibody; apoptosis; cancer; arthritis;  
XX KM autoimmune; ds.  
XX OS Homo sapiens.  
XX OS Synthetic.  
XX PN WO200073349-A1.  
XX PD 07-DEC-2000.  
XX PF 25-MAY-2000; 2000WO-US14599.  
XX PR 28-MAY-1999; 99US-0322875.  
XX PA (GETH ) GENENTECH INC.  
XX PI Ashkenazi AJ, Chuntharapai A, Dodge KH, Kim KJ;  
XX DR WPI; 2001-041145/05.  
XX DR  
XX PT Novel anti-death receptor 4 antibodies useful for treating cancer and  
XX PT immune related disorders such as rheumatoid arthritis, sjogren's  
XX PT syndrome, Grave's disease and diabetes mellitus -  
XX PS Disclosure; Fig 18; 126pp; English.  
XX CC The present invention relates to an anti-death receptor 4 (DR4)  
XX CC antibody. The antibodies of the invention are useful for inducing  
XX CC apoptosis in mammalian cancer cells such as colon cancer cells and  
XX CC for treating an immune-related disease in a mammal such as arthritis  
XX CC and autoimmune disease.  
XX SQ Sequence 1431 BP; 273 A; 367 C; 438 G; 350 T; 3 other;

Query Match 80.7%; Score 271; DB 22; Length 1431;  
Best Local Similarity 96.5%; Pred. No. 2.9e-73;  
Matches 277; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CTCGAGAGTCAGAGACCTGGCTGCTGTCGCCCTCACAGAGCCGTCATCATCTTGACCT 60  
Db 1365 CTCAGAGAGTCAGAGACCTGGCTGCTGTCGCCCTCACAGAGCCGTCATCATCTTGACCT 1306  
QY 61 GTCCTGGGTTTTCATTAAACAGCTATGTTACACTGGGTTGGCCAGCTCCAGGAAG 120

Db 1205 GTCCTGGGTTTTCATTAAACAGCTATGTTACACTGGGTTGGCCAGCTCCAGGAAG 1246  
QY 121 GGTCTGGAGTGGCTGGGAGTCAATTTGGGCTGTGTGAAGCACAATTTATATTCGGCTTC 180  
Db 1245 GGTCTGGAGTGGCTGGGAGTCAATTTGGGCTGTGTGAAGCACAATTTAAATTCGGCTTC 1186  
QY 181 ATGTCCAGACTGAACATCAACAGACAGCAATTCACAAGCCAAATTTCTTAAAAATGAAAC 240  
Db 1185 ATGTCCAGACTGAGCATCAGCAAAAGACACTCCAAAGACCAGTTTCTTAAAAATGAAAC 1126  
QY 241 AGTCTGAAACTGATGACAGCCATCTACTGTGCTCCAGAGAGG 287  
Db 1125 AGTCTGAAACTGATGACAGCCATCTACTGTGCTCCAGAGAGG 1079

RESULT 11  
AAA38902  
ID AAA38902 standard; DNA: 357 BP.

XX AAA38902;  
AC  
XX 29-AUG-2000 (first entry)  
XX DT  
XX DE 260F9 hybridoma VH domain encoding DNA SEQ ID NO:13.  
XX KW Antigen binding site; immunoglobulin; cancer antigen; immunological;  
XX KM antibody; tumour; human; mucin; cancer; cytostatic; hybridoma;  
XX KW specific binding assay; affinity purification; drug targeting;  
XX KM toxin targeting; imaging; genetic; therapeutic; ss.  
XX OS Homo sapiens.  
XX PN US6054561-A.  
XX PD 25-APR-2000.  
XX PF 07-JUN-1995; 95US-0483749.  
XX PR 21-MAR-1986; 86US-0842476.  
XX PR 08-MAY-1988; 88US-0190778.  
XX PR 08-FEB-1984; 84US-0577976.  
XX PR 11-JAN-1985; 85US-0690750.  
XX PR 11-AUG-1994; 94US-0288981.  
XX PA (CHIR ) CHIRON CORP.  
XX PI  
XX PI Ring DB;  
XX DR WPI; 2000-338508/29.  
XX DR P-PSDB; AAY90818.  
XX XX  
XX PT Monoclonal antibody capable of binding to human breast cancer antigen  
XX PT useful for affinity purification, drug or toxin targeting, imaging, and  
XX PT treating cancer -  
XX PS Disclosure; Fig 7; 57pp; English.  
XX CC The present invention describes a monoclonal antibody (MAb) (I) that  
XX CC binds to a human breast cancer antigen that is also bound by MAb 454C11  
XX CC and 520C9 (produced by hybridoma ATCC HB8484 and HB8696, respectively). Also  
XX CC described is a hybridoma that produces (I). (I) is useful in specific  
XX CC binding assays, affinity purification, drug or toxin targeting, imaging,  
XX CC and genetic or immunological therapeutics for various cancers. The  
XX CC present sequence encodes a VH domain derived from a 260F9 hybridoma,  
XX CC which is used in the exemplification of the present invention.  
XX SQ Sequence 357 BP; 89 A; 89 C; 93 G; 86 T; 0 other;

Query Match 80.4%; Score 270.2; DB 21; Length 357;  
Best Local Similarity 89.4%; Pred. No. 2.9e-73;  
Matches 303; Conservative 0; Mismatches 33; Indels 3; Gaps 1;

xx	Sequence	304 BP; 79 A; 74 C; 81 G; 70 T; 0 other;
SQ		
	Query Match	80.3%; Score 269.8; DB 20; Length 304;
	Best Local Similarity	95.8%; Pred. No. 3.7e-73;
	Matches 277; Conservative	0; Mismatches 12; Indels 0; Gaps 0;
OY	1 CTGCAGAGTCAGGACCTGGCTGGTGGCCCTTACAGAGCCTGTCCATCCTTGCACT	60
Db	10 ctgaagagatcagagacctgacctgctgctgagccctacaagagcctgctccatcacttgca	69
OY	61 GTCCTGGGTTTTCATTAAACGACGATGGTGTACACTGGGTTCCGACAGCTCCAGGAAG	120
Db	70 gctcttgagtttctatlaacagcctatgctgctacactggtgttcgcacagctccaggaag	129
OY	121 GGTCTGGAGTGCGCTGGAGTCATTTGGCTGGTGGAGCAGCAAAATTATTAATCGGCTTC	180
Db	130 ggtctggagtggtcgtgggagtaatatggtgctgtgtggaagacaataattatattcgtcttc	189
OY	181 ATGTCCAGACTGACATCATCAGAGACAATTCCAAGAGGCCAAATTTCTTAAAAATGAAC	240
Db	190 atgtccaaactatgagatcagcaagaacaacttccaagagagcaagtttctttaaanaatgaac	249

Feature	Location/Qualifiers
AAQ37462	standard; DNA; 669 BP.
AAQ37462:	
19-JUN-1993	(first entry)
VH_NQ2/12.4-Vk_NQ10/12.5	linked coding sequence #2.
primer:	human; immunoglobulin; Ig; variable region; VH; VL; Ck; JH; lymphocyte; vector; soluble; antibody; phage; linker; back; VH3; nested; in-cell PCR; cloning; polymorphic; TCR V; antiphenylloxazalone hydridoma; NQ2/12.4; NQ10/12.5; ss.
Synthetic.	
Key	
misc_RNA	Location/Qualifiers
	1..339
	/*tag= a
	/label= VH_NQ2/12.4
	1..23
	/*tag= b
	/label= NQ2/12.4_BACK
	38..58
	/*tag= c
	/label= NQ2/12.4BKNES
	complement (41..60)
	/*tag= d
	/label= NQ2HPRB
	complement (307..352)
	/*tag= e
	/label= MOVHlnk4
	340..345
	/*tag= f
	/note= "linker peptide"
	346..669
	/*tag= g
	/label= Vkappa_NQ2/12.4
	327..370
	/*tag= h
	/label= MOVkappaInk4
	complement (435..453)
	/*tag= i



```

FT      primer_bind          /label= NO2kappaPRB
FT                                         /complement (625..645)
FT                                         /*tag= j
FT                                         /label= MOKappa5FORNES
FT      primer_bind          /label= complement (647..669)
FT                                         /*tag= k
FT                                         /label= MOJkappa5FOR2
XX
XX      WO9303151-A.
XX
XX      18-FEB-1993.
XX
XX      10-AUG-1992;       92WO-GB01483.
XX
XX      10-AUG-1991;       91GB-0017352.
XX      PR
XX      11-JUN-1992;       92GB-0012419.
XX
XX      (MEDIC-) MEDICAL RES COUNCIL.
PA
PI      Embleton MJ, Gorochov G, Jones PT, Winter GP;
DR      WI; 1993-076508/09.
DR      P-PSDB; AAR32843.
XX
XX      Treatment of cell populations, partic. hybridomas - to link
PT      together copies of 2 or more non-contiguous DNA sequences to
PT      facilitate analysis
XX
XX      Disclosure; Fig 5; 72pp: English.
XX
XX      The sequences given in AA037459-62 show the nucleotide sequences of
CC      mature heavy chain VH domains and the VK light chain genes of the
CC      antiphenylloxazalone hydridomas NQ2/12.4 and NQ10/12.5 which have been
CC      linked via a linker peptide by in-cell PCR. The cDNA was synthesised
CC      using forward primers annealing to the Ck gene and the JH segment,
CC      followed by assembly with linker primers, VH back primers based on the
CC      VH3 leader sequence and a forward CK primer nested in respect to the
CC      primer used for cDNA. The assembled product within the cells is then
CC      amplified with nested primers annealing to the 5' end of the VH gene
CC      and the 3' end of the JK segment. In-cell PCR may be used to
CC      determine gene linkage analysis, particularly for the cloning of gene
CC      combinations that are polymorphic within a population of cells, such
CC      as the rearranged genes for Ig or TCR V regions.
XX
XX      Sequence 669 BP; 165 A; 178 C; 174 G; 152 T; 0 other;
SQ

```

AAQ37459	standard; DNA; 708 BP.
ID	AAQ37459
AC	AAQ37459;
XX	
DT	19-JUN-1993 (first entry)
XX	
DE	VH NQ2/12.4-VK NQ10/12.5 linked coding sequence.
XX	
KW	Primer: human; immunoglobulin; Ig; variable region; VH; VL; Ck; JH;
KW	lymphocyte; vector; soluble; antibody; phage; linker; back; VH3;
KW	nested; in-cell PCR; cloning; polymorphic; TCR V; antiphenylloxazolone;
KW	hybridoma; NQ2/12.4; NQ10/12.5; ss.
XX	
OS	Synthetic.
XX	
XX	
PH	Key
FT	1..339
FT	/*tag= a
FT	/label= VH_NQ2/12.4
FT	1..23
FT	/*tag= b
FT	/label= NQ2/12.4_back
FT	38..58
FT	/*tag= c
FT	/label= NQ2/12.4BKNES
FT	complement (41..60)
FT	/*tag= d
FT	/label= NQ2HPRB
FT	complement (322..371)
FT	/*tag= e
FT	/label= MOVhlnk3
FT	340..384
FT	/*tag= f
FT	/note= "Linker Peptide"
FT	385..708
FT	/*tag= g
FT	/label= VK_NQ10/12.5
FT	353..410
FT	/*tag= h
FT	/label= MOVkappaLnk3
FT	complement (474..492)
FT	/*tag= i
FT	/label= NQ2KPRB
FT	complement (664..684)
FT	/*tag= j
FT	/label= MOKkappa5FORNES
FT	complement (686..708)
FT	/*tag= k
FT	/label= MOKkappa5FOR2
XX	
PN	W09303151-A.
XX	
PD	18-FEB-1993.
XX	
PF	10-AUG-1992; 92MO-GB01483.
XX	
PR	10-AUG-1991; 91GB-0017352.
PR	11-JUN-1992; 92GB-0012419.
XX	
PA	(MED1-) MEDICAL RES COUNCIL.
XX	
PI	Embleton MJ, Gorochoy G, Jones PT, Winter GP;
XX	WPI: 1993-076508/09.
DR	P-PSDB: AAR32840.
XX	
XX	Treatment of cell populations, partic. hybridomas - to link
PT	together copies of 2 or more non-contiguous DNA sequences to
XX	facilitate analysis
PS	Disclosure; Fig 2; 72pp; English.

CC The sequences given in AA037459-62 show the nucleotide sequences of  
CC mature heavy chain VH domains and the VK light chain genes of the  
CC anti-phenylloxazole hydriodomas NQ2/12.4 and NQ10/12.5 which have been  
CC linked via a linker peptide by In-cell PCR. The cDNA was synthesised  
CC using forward primers annealing to the CK gene and the JH segment,  
CC followed by assembly with linker primers, VH back primers based on the  
CC VH3 leader sequence and a forward CK primer nested in respect to the  
CC primer used for cDNA. The assembled product within the cells is then  
CC amplified with nested primers annealing to the 5' end of the VH gene  
CC and the 3' end of the JK segment. In-cell PCR may be used to  
CC determine gene linkage analysis, particularly for the cloning of gene  
CC combinations that are polymorphic within a population of cells, such  
CC as the rearranged genes for Ig or TCR V regions.  
XX  
SQ Sequence 708 BP; 166 A; 184 C; 198 G; 160 T; 0 other;

Query Match 80.2%; Score 269.6; DB 14; Length 708;  
Best Local Similarity 96.8%; Pred. No. 5.9e-73;  
Matches 275; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGAGCTGTCCATCTGCACT 60  
DB 10 ctgaagaggtcaagaccctggcctggtggccctcaccagagcctgtccatcactgcaact 69  
QY 61 GTCCTGGGTTTTCATTAAACAGCTATGCTGACCTGGTTCGCCAGCTCCAGGAAG 120  
DB 70 gctctgggtttcatttaaccagctatggtgtacactggttcgcagcctccagaaga 129  
QY 121 GGTCTGAGTGGCTGGGCTGATTTGGCTGTGGAACACAAATTATTAATTTGGCTCTC 180  
DB 130 ggtctggagtggtcgggagtaataatggcctggtaagacacaataatcggctctc 189  
QY 181 ATGTCACAGCTGAACATCAACAGACAAATTCACAGAGCCAAATTTCTTAAATGAAC 240  
DB 190 atgtccagactgagatcagcaagaacaccccaagagcgaagtttctttaaataatgaaac 249  
QY 241 AGTCTGCAAACTGATGACACAGCCATCTACTACTGTGCCAGCA 284  
DB 250 agctcgcaactgacgacacagccatgtactactgtgccaaga 293

RESULT 15  
AAZ31383  
ID AAZ31383 standard; DNA; 291 BP.  
XX  
AC AAZ31383;  
XX  
DN 07-FEB-2000 (first entry)  
XX

DE MUSIGHAD mouse Ig heavy chain mRNA V region, partial CDS.

XX Monoclonal antibody; Mab; 1A7; GD2; immune response; melanoma;  
KW neuroblastoma; glioma; soft tissue carcinoma; small cell carcinoma;  
XX tumor-associated antigen; ss.

XX Mus sp.

XX US5977316-A.

XX 02-NOV-1999.

XX 16-JAN-1996; 96US-0591196.

XX 17-JAN-1995; 95US-0372676.

XX (KENT ) UNIV KENTUCKY.

XX Foon KA, Chatterjee SK, Chatterjee M;

XX WPI; 1999-619711/53.

XX Monoclonal antibody 1A7 which elicits an anti-GD2 immunological

PT response, useful for the development of products for the detection and  
PT treatment of cancers -  
XX  
XX  
PS Disclosure; Fig 13C; 74pp; English.

CC The invention provides a monoclonal antibody (MAB) designated 1A7, which  
CC elicits an anti-GD2 (tumor-associated antigen) immunological response in  
CC humans. Mab 1A7 has defined light and heavy chain variable region  
CC sequences. The Mab 1A7 and polypeptides can be used for eliciting an  
CC anti-GD2 immune response. The polypeptides can also be used for  
CC detecting or purifying anti-GD2 antibody. The products can be used for  
CC treating GD2-associated diseases, e.g. melanoma, neuroblastoma, glioma,  
CC soft tissue carcinoma, and small cell carcinoma. They can be used for  
CC palliating the disease or for reducing the risk of recurrence. Sequences  
CC AAZ31373-90 represent sequences that closely matched Mab 1A7 heavy chain  
CC variable region encoding sequence.  
XX  
SQ Sequence 291 BP; 76 A; 71 C; 77 G; 67 T; 0 other;

Query Match 79.6%; Score 267.6; DB 20; Length 291;  
Best Local Similarity 96.8%; Pred. No. 1.7e-72;  
Matches 273; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CTGCAGAGTCAAGACCTGGCTGGTGGCCCTCAGAGAGCTGTCCATCTGCACT 60  
DB 10 ctgaagaggtcaagaccctggcctggtggccctcaccagagcctgtccatcactgcaact 69  
QY 61 GTCCTGGGTTTTCATTAAACAGCTATGCTGACCTGGTTCGCCAGCTCCAGGAAG 120  
DB 70 gctctgggtttcatttaaccagctatggtgtacactggttcgcagcctccagaaga 129  
QY 121 GGTCTGAGTGGCTGGGCTGATTTGGCTGTGGAACACAAATTATTAATTTGGCTCTC 180  
DB 130 ggtctggagtggtcgggagtaataatggcctggtaagacacaataatcggctctc 189  
QY 181 ATGTCACAGCTGAACATCAACAGACAAATTCACAGAGCCAAATTTCTTAAATGAAC 240  
DB 190 atgtccagactgagatcagcaagaacaccccaagagcgaagtttctttaaataatgaaac 249  
QY 241 AGTCTGCAAACTGATGACACAGCCATCTACTACTGTGCCAGCA 282  
DB 250 agctcgcaactgacgacacagccatgtactactgtgccaaga 291

Search completed: August 20, 2002, 09:36:03  
Job time: 10478 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 08:55:02 ; Search time 2643.23 Seconds  
(without alignments)  
1715.695 Million cell updates/sec

Title: US-09-824-286-6

Perfect score: 336  
Sequence: 1 CTCGAGGAGTCGACCTGTG.....GCCAAGGAGCAGCGTCAAC 336

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	270.2	80.4	840	10	BI690298
2	249	74.1	691	10	BF578188
3	238.2	70.9	671	10	BF182141
4	236.8	70.5	930	10	BF579344
5	226.4	67.4	685	12	BH275985
6	225.2	67.0	520	9	AW917371
7	219	65.2	637	10	BF531263
8	210.8	62.7	359	9	BB870665
9	198.6	59.1	425	10	BF578521
10	195	58.0	694	10	BF134274
11	174.4	51.9	507	10	BE845785
12	172.8	51.4	476	10	BE482206
13	172.8	51.4	496	10	BE485461
14	172.6	51.4	472	10	BE482203
15	171.4	51.0	414	10	BE478938
16	171.2	51.0	489	10	BE476735
17	170.8	50.8	511	10	BE485545

18	170.6	50.8	548	10	BG690980
19	170.2	50.7	558	10	BE845893
20	170	50.6	504	10	BE485128
21	169	50.3	523	10	BE483736
22	168.8	50.2	373	10	BE487845
23	168.6	50.2	413	10	BE481411
24	168.2	50.1	420	10	BF230480
25	167.8	49.9	497	10	BE486006
26	167.8	49.9	547	10	BG692467
27	167	49.7	500	10	BG692280
28	167	49.7	530	10	BG692277
29	167	49.7	549	10	BE845758
30	166.6	49.6	502	10	BE485907
31	166.4	49.5	505	10	BE589545
32	166.2	49.5	513	10	BG692944
33	166.2	49.5	602	10	BE588767
34	165.8	49.3	382	10	BE485343
35	165.4	49.2	501	10	BF230469
36	165.4	49.2	509	10	BE480721
37	164.8	49.0	513	10	BE483381
38	164.6	49.0	390	10	BE482316
39	164.6	49.0	412	10	BE482746
40	164.4	48.9	524	10	BE476249
41	164.2	48.9	376	10	BE486964
42	164	48.8	405	10	BE484472
43	164	48.8	511	10	BE478595
44	164	48.8	598	10	BE588685
45	163.6	48.7	381	10	BE488145

## ALIGNMENTS

RESULT 1  
BI690298 840 bp mRNA Linear EST 18-SEP-2001  
LOCUS 603309721.F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5345741 5',  
DEFINITION mRNA sequence.  
ACCESSION BI690298  
VERSION BI690298.1 GI:15652927  
KEYWORDS EST,  
SOURCE house mouse,  
ORGANISM Mus musculus  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS 1 (bases 1 to 840)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM1878 row: 1 column: 06  
High quality sequence stop: 828.  
Location/Qualifiers  
1. 840  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5345741"  
/clone\_lib="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT."

Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH\*

BASE COUNT 196 a 237 c 214 g 193 t

ORIGIN

Query Match 80.4%; Score 270.2; DB 10; Length 840;  
Best Local Similarity 89.4%; Pred. No. 5.6e-72;  
Matches 303; Conservative 0; Mismatches 33; Indels 3; Gaps 1;

QY 1 CTGACGAGTACAGACCTGGCCGTGGCCCTCAGACAGCTGTCCATCATCTGCAC 60  
DB 85 CTGAAGAGTACAGACCTGGCCGTGGCCCTCAGACAGCTGTCCATCATCTGCAC 144  
QY 61 GTCCTGGGTTTCATTAAACAGCTATGTTAGACTGGTTCGCCACCTCCAGAAAG 120  
DB 145 GTCCTGGGTTTCATTAAACAGCTATGTTAGACTGGTTCGCCACCTCCAGAAAG 204  
QY 121 GGTCTGAGTGGCTGGGAGTATTTGGCTGTGGAAACAAATTAATTAATTCGGCTTC 180  
DB 205 GGTCTGAGTGGCTGGGAGTATTTGGCTGTGGAAACAAATTAATTAATTCAGCTTC 264  
QY 181 ATGTCCAGACTGACATCAACAGACAAATTCAGAGCCAAATTTCTTAAAAATGAAC 240  
DB 265 AAATCCAGACTGACATCAACAGACAAATTCAGAGCCAAATTTCTTAAAAATGAAC 324  
QY 241 AGTTCGAACTGATGACACAGCCATCTACTGTGTGCCAGAGAGGTTT---CTACGGTA 297  
DB 325 AGTTCGAACTGATGACACAGCCATCTACTGTGTGCCAGAGAGGTTTACTAACGGAG 384  
QY 298 GATTCGACTGACTGGGGCCAGGGGACACGGGTCCAC 336  
DB 385 GGTCTGAGTGGCTGGGAGTATTTGGCTGTGGAAACAAATTAATTAATTCAGCTTC 423

RESULT 2  
BF578188 691 bp mRNA linear EST 12-DEC-2000  
LOCUS 602094691.F1 NCL\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4209022 5',  
DEFINITION mRNA sequence.  
ACCESSION BF578188  
VERSION BF578188.1 GI:11651900  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapds-femail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM9774 row: e column: 23  
High quality sequence stop: 690.

FEATURES  
source location/Qualifiers  
1..691  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4209022"  
/clone\_1lb="NCI\_CGAP\_Co24"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: oligo dt.  
Average insert size 1.6 kb. Constructed by Life

Technologies. Note: this is a NCL\_CGAP Library."

BASE COUNT 174 a 195 c 171 g 151 t

ORIGIN

Query Match 74.1%; Score 249; DB 10; Length 691;  
Best Local Similarity 86.3%; Pred. No. 1.7e-65;  
Matches 289; Conservative 0; Mismatches 40; Indels 6; Gaps 1;

QY 1 CTGACGAGTACAGACCTGGCCGTGGCCCTCAGACAGCTGTCCATCATCTGCAC 60  
DB 82 CTGAAGAGTACAGACCTGGCCGTGGCCCTCAGACAGCTGTCCATCATCTGCAC 141  
QY 61 GTCCTGGGTTTCATTAAACAGCTATGTTAGACTGGTTCGCCACCTCCAGAAAG 120  
DB 142 GTCCTGGGTTTCATTAAACAGCTATGTTAGACTGGTTCGCCACCTCCAGAAAG 201  
QY 121 GGTCTGAGTGGCTGGGAGTATTTGGCTGTGGAAACAAATTAATTAATTCGGCTTC 180  
DB 202 GGTCTGAGTGGCTGGGAGTATTTGGCTGTGGAAACAAATTAATTAATTCAGCTTC 261  
QY 181 ATGTCCAGACTGACATCAACAGACAAATTCAGAGCCAAATTTCTTAAAAATGAAC 240  
DB 262 AAATCCAGACTGACATCAACAGACAAATTCAGAGCCAAATTTCTTAAAAATGAAC 321  
QY 241 AGTTCGAACTGATGACACAGCCATCTACTGTGTGCCAGAGAGGTTTCTAACGGTAGAT 300  
DB 322 AGTTCGAACTGATGACACAGCCATCTACTGTGTGCCAGAGAGGTTTCTAACGGTAGAT 376  
QY 301 TCTATGACTGACTGGGGCCAGGGGACACGGGTCCAC 335  
DB 377 -ACTTGTGACTGAGGGCCAGGGGACACCTCTCAC 410

RESULT 3  
BF182141 671 bp mRNA linear EST 31-OCT-2000  
LOCUS 601804682.F1 NCL\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4035429 5',  
DEFINITION mRNA sequence.  
ACCESSION BF182141  
VERSION BF182141.1 GI:11060283  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgapds-femail.nih.gov  
Tissue Procurement: Lotmar Hennighausen Ph.D., Robin Humphreys  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM9309 row: d column: 22  
High quality sequence stop: 669.

FEATURES  
source location/Qualifiers  
1..671  
/organism="Mus musculus"  
/strain="C57/B6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4035429"  
/clone\_1lb="NCI\_CGAP\_Mam5"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: oligo dt.

Library constructed by Life Technologies. Investigators  
providing samples: Lochar Hennighausen/Robin Humphreys,  
NIH

BASE COUNT 173 a 177 c 164 g 157 t  
ORIGIN

Query Match 70.9%; Score 238.2; DB 10; Length 671;  
Best Local Similarity 83.5%; Pred. No. 3,4e-62;  
Matches 283; Conservative 0; Mismatches 53; Indels 3; Gaps 1;

QY 1 CTGCAGAGACTCAGAGACCTGGCTGGTGGCCCTCAGAGACCTGTCCATCTGCACT 60  
DB 67 CTGAGAGACTCAGAGACCTGGCTGGTGGCCCTCAGAGACCTGTCCATCTGCACT 126  
QY 61 GTCCTGGGTTTTCATTACACGCTATGGTACACTGGTTCGCCAGCTCCAGGAAG 120  
DB 127 GTCTCGAGTTCTCATTAACGCTATGATTAACCTGTGATGCGCAACACAGGAAG 186  
QY 121 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAAGCACAATTAATTAATTCGCTTC 180  
DB 187 GGTCTGAGTGGCTGGAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTCGCTTC 246  
QY 181 ATGTCCAGACTGACATCAACAGACAAATTCACAGACCAATTTCTTTAAATGAAC 240  
DB 247 ATGTCCAGACTGACATCAACAGACAAATTCACAGACCAATTTCTTTAAATGAAC 306  
QY 241 AGTCTGCAACATGATGACACGCTATCTACTACTGTGCCAGA---GAGGTTCTACGGTA 237  
DB 307 AGTCTGCAACATGATGACACGCTATCTACTACTGTGCCAGA---GAGGTTCTACGGTA 366  
QY 298 GATTCATGACTACTGAGGCGCAAGGACACGCTACCC 336  
DB 367 TATGTATGACTACTGAGGCGTCAAGAACCTCAGTCAAC 405

RESULT 4  
BF579344 930 bp mRNA linear EST 12-DEC-2000  
LOCUS BF579344  
DEFINITION 602095631P1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4215751 5',  
mRNA sequence.  
ACCESSION BF579344  
VERSION BF579344.1 GI:11653056  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 930)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM9791 row: n column: 08  
High quality sequence stop: 633.  
Location/Qualifiers  
1..930  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4215751"  
/clone\_lib="NCI\_CGAP\_Co24"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NciI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.6 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP library."

BASE COUNT 240 a 237 c 249 g 204 t  
ORIGIN

Query Match 70.5%; Score 236.8; DB 10; Length 930;  
Best Local Similarity 81.5%; Pred. No. 1e-61;  
Matches 274; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 CTGCAGAGACTCAGAGACCTGGCTGGTGGCCCTCAGAGACCTGTCCATCTGCACT 60  
DB 84 CTGAGAGACTCAGAGACCTGGCTGGTGGCCCTCAGAGACCTGTCCATCTGCACT 143  
QY 61 GTCCTGGGTTTTCATTACACGCTATGGTACACTGGTTCGCCAGCTCCAGGAAG 120  
DB 144 GTCTCGAGTTCTCATTAACGCTATGATTAACCTGTGATGCGCAACACAGGAAG 203  
QY 121 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAAGCACAATTAATTAATTCGCTTC 180  
DB 204 GGTCTGAGTGGCTGGAGTCAATTTGGCTGGTGAAGCACAATTAATTAATTCGCTTC 263  
QY 181 ATGTCCAGACTGACATCAACAGACAAATTCACAGACCAATTTCTTTAAATGAAC 240  
DB 264 AAATCCAGACTGACATCAACAGACAAATTCACAGACCAATTTCTTTAAATGAAT 323  
QY 241 AGTCTGCAACATGATGACACGCTATCTACTACTGTGCCAGAAGGTTCTACGGTAAT 300  
DB 324 AGTCTGCAACATGATGACACGCTATCTACTACTGTGCCAGAAGGTTCTACGGTAAT 383  
QY 301 TCTATGACTACTGAGGCGCAAGGACACGCTACCC 336  
DB 384 ACTATGACTACTGAGGCGTCAAGAACCTCAGTCAAC 419

RESULT 5  
BH275985 685 bp DNA linear GSS 30-NOV-2001  
LOCUS BH275985/c  
DEFINITION CH230-96J24, TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-96J24, DNA sequence.  
ACCESSION BH275985  
VERSION BH275985.1 GI:17188387  
KEYWORDS GSS.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 685)  
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,  
A., Gebregregorys,E., Overton,L., Russell,D., Chen,D., Riggs,F., de  
Jong,P. and Fraser,C.M.  
Rat BAC End Sequences from Library CHORI-230 EcORI segment  
Unpublished (1999)  
Other GSSs: CH230-96J24..TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhao@tigr.org](mailto:szhao@tigr.org)  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdejong@tigr.org](mailto:pdejong@tigr.org)).  
Clones may be purchased from BACPAC Resources  
([http://www.chori.org/bacpac/or\\_eing\\_information.htm](http://www.chori.org/bacpac/or_eing_information.htm)). BAC end  
page: [http://www.tigr.org/tadb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html)  
Plate: 96 row: J column: 24  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1..685

FEATURES  
source

```
/organism="Rattus norvegicus"
/strain="BN/SSNhsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-96J24"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/Note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SSNhsd/MCw) BAC library produced by
Pleier de Jong"

BASE COUNT      151 a      167 c      174 g      193 t
ORIGIN

Query Match      67.4%; Score 226.4; DB 12; Length 685;
Best Local Similarity 87.3%; Pred. No. 1.4e-58;
Matches 248; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 CTGCAGAGTCAGAGACCTGGCCCTGTGGCCCTCAGACAGCCCTGTCATCCTTGCACCT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 CTGAAGAGTCAGAGACCTGGCTGTGGCTGACGCCCTCAGACAGCCCTGTCATCCTTGCACCT 345

QY 61 GTCCTGGGTTTCATTAACACAGCTATGTTGACCTGGGTTGGCAGCCCTCCAGGAAG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 GTCCTGGGTTTCATTAACACAGCTATGTTGACCTGGGTTGGCAGCCCTCCAGGAAG 285

QY 121 GGTCTGGAGTGGCTGAGTCATTTGGCTGGTGGAGACACAATTAATTCGGCTCTC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 GGTCTGGAGTGGATTCAGCAATATGATGGTGGAGACACAATTAATTCAGCTCTC 225

QY 181 ATGTCCAGACTGAGACATCAACAGACAAATTCCTTAATTAATTAATTAATTAAT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 AAATCCGACTGACATCAGACAGGACACCTCCAGAGCCAAAGTTTCTTAATAATGAAC 165

QY 241 AGTCTGCAACTGATGACACAGCCATCTACTACTGTGCCAGAGA 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 AGTCTGAAACTGAAACACAGCATATGATCTACTACTGTGCCAGAGA 121

RESULT 6
AM917371      520 bp      mRNA      linear      EST 25-MAY-2000
LOCUS      EST348675 Rat gene index, normalized rat, norvegicus, Bento Soares
DEFINITION      Rattus norvegicus CDNA clone RGIED01 5' end, mRNA sequence.
ACCESSION      AM917371
VERSION      AM917371.1 GI:8083120
KEYWORDS      EST.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 520)
Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Queckenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information
Seq primer: M13 Reverse.
Location/Qualifiers
1..520
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RGIED01"
/clone_lib="Rat gene index, normalized rat, norvegicus,"
```

```
Bento Soares"
/tissue_type="mix - brain, ovary, placenta, kidney, lung,
liver, embryo, heart, muscle, spleen"
/lab_host="SOLR"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Estimated insert size approx.1 kb"

BASE COUNT      122 a      135 c      137 g      125 t
ORIGIN

Query Match      67.0%; Score 225.2; DB 9; Length 520;
Best Local Similarity 79.4%; Pred. No. 3e-58;
Matches 266; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 CTGCAGAGTCAGAGACCTGGCCCTGTGGCCCTCAGACAGCCCTGTCATCCTTGCACCT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 CTGATGGAGTCAGAGACCTGGCTGTGGAGGCCCTCAGAGACCCCTGTCCTCCTGATTT 136

QY 61 GTCCTGGGTTTCATTAACACAGCTATGTTGACCTGGGTTGGCAGCCCTCCAGGAAG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 GTCCTGGGTTTCATTAACACAGCTATGTTGACCTGGGTTGGCAGCCCTCCAGGAAG 196

QY 121 GGTCTGGAGTGGCTGAGTCATTTGGCTGGTGGAGACACAATTAATTCGGCTCTC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 GGTCTGGAGTGGATGGAGTATGTGGAGTGGAGACACAATTAATTCAGCTCTC 256

QY 181 ATGTCCAGACTGAGACATCAACAGACAAATTCCTTAATTAATTAATTAATTAAT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 AAATCCGACTGACATCACCAGGAGACCTCCAAAGCAAGTTTCTTAATAATGAAC 316

QY 241 AGTCTGCAACTGATGACACAGCCATCTACTACTGTGCCAGAGAGGTTCTACGTTAGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 AGTCTGCAACTGATGAGACACCACTTACTACTGTGCCAGAGGTTCTACGTTAGAT 376

QY 301 TCTATGAGCTACTGGGGCCAGGAGCCAGGTCAC 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 TACTTTGATTACTGGGGCCAGGAGTCAATGTCAC 411

RESULT 7
BF531263      637 bp      mRNA      linear      EST 11-DEC-2000
LOCUS      BF531263
DEFINITION      60209108.F1 NCL_CGAP_Co24 Mus musculus CDNA clone IMAGE:4205516 5',
mRNA sequence.
ACCESSION      BF531263
VERSION      BF531263.1 GI:11618626
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 637)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM9765 row: c column: 21
High quality sequence stop: 635.
Location/Qualifiers
1..637
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4205516"
/clone_lib="NCL_CGAP_Co24"
```

URL:<http://genome.gsc.riken.go.jp/Carninci,P.,Shibata,Y.,Hayata,N.,Sugahara,Y.,Shibata,K.,Itoh,M.,Konno,H.,Okazaki,Y.,Matsutani,M.,andHayashizaki,Y.>

REFERENCE  
AUTHORS  
TITLE  
1 (bases 1 to 425)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)







JOURNAL  
COMMENT Unpublished (2000)  
Contact: Sonstegard TS  
USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8414  
Email: tads@ps1.barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -minscore 18  
and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 133 row: J column: 4  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. .496  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_id="BARC\_5BOV"  
/rissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;  
Library made from pooled mRNA isolated from mammary  
tissues at eight physiological, developmental, and disease  
states."  
BASE COUNT 108 a 151 c 138 g 99 t  
ORIGIN

Query Match 51.4%; Score 172.8; DB 10; Length 496;  
Best Local Similarity 75.0%; Pred. No. 3.5e-42;  
Matches 216; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 CTGCAGAGTACAGACCTGGCTGGGCGCCCTCAGACAGCCTGTCATCTTGCACT 60  
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 87 CTGGGGAGTGGGGCCCGACGCTGTGTAAGCCCTCAGACAGCCTGTCATCTTGCACT 146  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 61 GTCTGTGGGTTTCATTAAACCAAGCTATGCTGTACACTGGGTTCCGACCTCCAGAAAG 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 147 GTCTGTGATTCTCATTTAAAGCGGGAGATGTAGGCTGGGTCGCCAGAGCTCCAGAAAG 206  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 121 GGCTGGAGTGGGCTGGGAGTATTGGGCTGGTGAACCAAAATTAAATTCGGCTTC 180  
| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 207 GCGCTGGAGTGGCTGAGGCCATTAAAGGCTGTGGAACACAGATATACCCAGCCCTG 266  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 181 ATGTCCAGACTGAACATCAACAGACAGCAATTCACAAGACCAAAATTTCTTAAAAATGAAC 240  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 267 AAATCCCGACTCAGCATTCACCAAGACAACTCCAAAGACCAAGTCTCTATCACTGAGC 326  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 241 AGCTGCAAACTGATGACACAGCAGCATCTACTACTGTGGCAGAGAGGTT 288  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 327 AGCGTGACTACTGAAGACACGCGCCACATCACTACTGTGGAAGATGTT 374  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14  
BE482203 472 bp mRNA linear EST 28-AUG-2000  
LOCUS BE482203  
DEFINITION 167903 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE482203  
VERSION BE482203.1 GI:9601736  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 472)  
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and  
Wells,K.D.  
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary  
gland cDNA library  
JOURNAL Unpublished (2000)  
COMMENT

JOURNAL  
COMMENT Contact: Sonstegard TS  
USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8414  
Email: tads@ps1.barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -minscore 18  
and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 22 row: N column: 23  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. .472  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_id="BARC\_5BOV"  
/rissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;  
Library made from pooled mRNA isolated from mammary  
tissues at eight physiological, developmental, and disease  
states."  
BASE COUNT 108 a 126 c 132 g 106 t  
ORIGIN

Query Match 51.4%; Score 172.6; DB 10; Length 472;  
Best Local Similarity 74.6%; Pred. No. 3.9e-42;  
Matches 217; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CTGCAGAGTACAGACCTGGCTGGGCGCCCTCAGACAGCCTGTCATCTTGCACT 60  
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 160 CTGGGGAGTGGGGCCCGACGCTGTGTAAGCCCTCAGACAGCCTGTCATCTTGCACT 219  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 61 GTCTGTGGGTTTCATTAAACCAAGCTATGCTGTACACTGGGTTCCGACCTCCAGAAAG 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 220 GTCTGTGATTCTCATTTAAAGCGAGAAATATGTAGGCTGGGTCGCCAGAGCTCCAGAAAG 279  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 121 GGCTGGAGTGGGCTGGGAGTATTGGGCTGGTGAACCAAAATTAAATTCGGCTTC 180  
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 280 GCGCTGGAGTGGCTGAGGCCATTATATCTGTGGAACACAGATATACCCAGCCCTG 339  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 181 ATGTCCAGACTGAACATCAACAGACAGCAATTCACAAGACCAAAATTTCTTAAAAATGAAC 240  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 340 AAATCCCGACTCAGCATTCACCAAGACAGCTCCAAAGACCAAGTCTCTATCACTGAGC 399  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 241 AGCTGCAAACTGATGACACAGCAGCATCTACTACTGTGGCAGAGAGGTTCT 291  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 400 AGCGTGACTACTGAAGACACGCGCCACATCACTACTGTGGAAGATGTT 450  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15  
BE478938 414 bp mRNA linear EST 28-AUG-2000  
LOCUS BE478938  
DEFINITION 163420 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE478938  
VERSION BE478938.1 GI:9598471  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 414)  
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and  
Wells,K.D.  
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary  
gland cDNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Sonstegard TS

USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8414

Email: tad@psi.barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -minscore 18  
and -mismatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCGAGTCAGCAGC

Plate: 141 row: D column: 4

Seq primer: ATTAGTGACACTATAG.

## FEATURES

## source

1..414

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_11b="BARC 5B0V"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: PCMV SPORT6; Site.1: XbaI; Site.2: XhoI;

library made from pooled mRNA isolated from mammary

tissues at eight physiological, developmental, and disease

states."

BASE COUNT 88 a 125 c 116 g 85 t

## ORIGIN

## Query Match

Best Local Similarity 51.0%; Score 171.4; DB 10; Length 414;

Matches 214; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

OY 1 CTGCAGAGTCAGACCTGGCTGGGCGCCCTCACAGAGCCTGTCATCAGTTCGACT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 CTGCGGAGTCGGGCCCGCCAGCTGTAAGCCTCACAGACCTCTCCTCAGCTGCAGC 173
OY 61 GTCTGGGTTTTCATTAAACAGCTAGTGTACACTGGGTGCCAGCCTCCAGAAAG 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 GTCTGGATTCATGACAGCTAGTGTAACTGGGTCCGCCAGGCTCCAGGGAAG 233
OY 121 GGTCTGAGTGGCTGGAGTCATTGGGCTGGTGGGAAGCACAAATTATTTGGCTCTC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 GCGCTGAGTGGGTGGTGGCATCGGCATGAGAAACACATATCTAAGCAGCCCTG 293
OY 181 ATGTCCAGACTGAACATCAACAGACAAATTTCCAAAGCCAAATTTCTAAAAATGAAC 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 AAATCCCGGCTCAGCATCACCAGACAACTCCAAAGCCCAAGTCTCTGTCTCAGTGAGC 353
OY 241 AGTCTGCAAACTGATGACAGAGCCATCTACTAGTGGCCAGAGAG 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 AGCGTGACACTGAGGACAGCCACATCTACTAGTGCAGAGAGG 398

```

Search completed: August 20, 2002, 08:55:05  
Job time: 11780 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:26 ; Search time 94.22 seconds  
(without alignments)  
5.894 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_28\_32

Perfect score: 29

Sequence: 1 SYGVH 5

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	5	17	AA898945
2	29	100.0	83	22	AAU22812
3	29	100.0	83	22	AAU96118
4	29	100.0	107	13	AA828434
5	29	100.0	112	19	AAW31648
6	29	100.0	113	13	AA821268
7	29	100.0	115	18	AAW04595
8	29	100.0	116	11	AA807332
9	29	100.0	116	13	AA828287
10	29	100.0	116	17	AAW14450
11	29	100.0	116	17	AA898877

12	29	100.0	119	13	AA825728	Humanised VH regio
13	29	100.0	119	20	AA492918	Heavy chain variab
14	29	100.0	119	20	AA49808	K588 antibody heav
15	29	100.0	119	22	AA869650	Human Lay antibody
16	29	100.0	119	22	AA869675	Murine mIk-beta1 a
17	29	100.0	119	22	AA869676	Humanised mIk-beta
18	29	100.0	123	18	AAW07438	Anti-DNA antibody
19	29	100.0	127	11	AA806355	Peptide correspond
20	29	100.0	133	18	AAW10546	Wild type murine a
21	29	100.0	133	18	AAW10542	Humanised murine a
22	29	100.0	133	18	AAW10543	Humanised murine a
23	29	100.0	133	18	AAW10539	Humanised murine a
24	29	100.0	133	18	AAW10540	Humanised murine a
25	29	100.0	133	18	AAW10541	Humanised murine a
26	29	100.0	133	18	AAW10538	Humanised murine a
27	29	100.0	138	22	AA869656	Murine mIk-beta1 a
28	29	100.0	140	18	AAW22558	Murine anti-human
29	29	100.0	144	17	AA898944	Humanised anti-CD3
30	29	100.0	151	17	AA898941	Humanised anti-CD3
31	29	100.0	174	21	AA493723	The heavy chain of
32	29	100.0	188	22	AAE05352	Mouse secreted pro
33	29	100.0	219	15	AA856235	b66-118/h13-65/11-
34	29	100.0	222	14	AA822843	VH NQ2/12.4-VK NQ1
35	29	100.0	223	15	AA856236	166-111/112-65/h1-
36	29	100.0	235	14	AA832840	VH NQ2/12.4-VK NQ1
37	29	100.0	240	22	AAU33861	Staphylococcus aur
38	29	100.0	240	22	AAU36912	Staphylococcus aur
39	29	100.0	249	21	AA495527	Anti-CD38 antibody
40	29	100.0	260	21	AA495527	Anti-CD38 antibody
41	29	100.0	264	16	AA895570	WOW-1 Fab heavy ch
42	29	100.0	269	13	AA832569	Intracellular bind
43	29	100.0	440	19	AAW42391	Fusion protein enc
44	29	100.0	450	21	AA44991	Thermococcus celer
45	29	100.0	456	21	AA44992	M79scfv-interleuk1

#### ALIGNMENTS

RESULT 1	
AA898945	standard; Protein: 5 AA.
ID	AA898945
XX	
AC	AA898945:
XX	
DT	03-JAN-1997 (first entry)
XX	
DE	CDR of humanised anti-CD38 monoclonal antibody heavy chain.
XX	
KW	Monoclonal antibody; Mab; humanised; cancer; autoimmune disease;
KW	multiple myeloma; lymphoma; rheumatoid arthritis; CD38;
KW	complementary determining region; CDR; heavy chain; light chain;
KW	ss.
XX	
OS	Synthetic.
XX	
PN	WC9616990-A1.
XX	
PD	06-JUN-1996.
XX	
PF	28-NOV-1995; 95WO-GB02777.
XX	
PR	02-DEC-1994; 94GB-0024449.
XX	
PA	(WELL ) WELLCOME FOUND LTD.
XX	
PI	Ellis JH, Lewis AP;
XX	
DR	WPI; 1996-277724/28.
XX	
PT	Humanised monoclonal antibodies with donor framework residues 29 and
PT	78 - esp. against CD38, useful for treating cancer and auto-immune
PT	diseases



PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251969.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI: 2001-451929/48.  
XX N-PSDB; AAS40179.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
XX disorders related to the reproductive system including prostate cancer  
XX and also for testing and detection e.g. diagnosis -  
XX  
XX Claim 11; SEQ ID NO 331; 546bp; English.  
XX  
XX The invention relates to novel isolated human prostate cancer antigen  
XX polynucleotides (I) and polypeptides (II). (I) and (II) are useful for  
XX preventing, treating or ameliorating a medical condition when  
XX administered. (I), (II) and the antibody to (II) are useful for treating,  
XX preventing and/or prognosing disorders related to the reproductive  
XX system including prostate cancers; urinary disorders e.g. chronic  
XX nephritis; and blood-related disorders e.g. thrombosis. (II) can be used  
XX for testing and detection e.g. as a chromosomal marker and in forensics.  
XX (I) and the anti-(II) antibody can be used in testing and detection in  
XX immunoassays. AAU22702-AAU22913 represent the human prostate cancer  
XX antigen amino acid sequences, and related amino acid sequences of the  
XX invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at: ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 83 AA:

Query Match 100.0%; Score 29; DB 22; Length 83;

Best Local Similarity 100.0%; Pred. No. 58;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYGVH 5  
Db 16 sygvh 20  
RESULT 3  
ID AAM96118 standard; Protein: 83 AA.  
XX  
XX AAM96118;  
AC  
XX  
XX 21-NOV-2001 (first entry)  
DE  
XX Human reproductive system related antigen SEQ ID NO: 4776.  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX WO200155320-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01339.  
PE  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198122.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225271.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226281.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PR XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI: 2001-465570/50.  
 DR N-PSDB: AAL02088.  
 XX  
 PT Isolated nucleic acid molecule encoding a reproductive system antigen  
 PI is used in preventing, treating or ameliorating a medical condition -  
 XX  
 BS Claim 11; SEQ ID NO 4776; 1297bp + Sequence Listing; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a protein of the invention.  
 XX  
 SQ Sequence 83 AA;  
 QY 1 SYGVH 5  
 DB 16 sygvh 20  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 4  
 AAR28434  
 ID AAR28434 standard; Protein; 107 AA.  
 XX  
 AC AAR28434;  
 XX  
 DT 04-APR-1993 (first entry)  
 XX  
 DE Sequence of monoclonal antibody NM-2 VH domain.  
 XX  
 KW Antibody; PCR primer; variable heavy chain;  
 KW anti-mucin monoclonal antibody.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 28..37  
 FT Region /label= CDR 1  
 FT Region 47..62  
 FT Region /label= CDR 2  
 FT Region 95..103  
 FT Region /label= CDR 3  
 XX  
 PN W09218534-A.



XX 29-OCT-1992.  
PD 92WO-GB00746.  
XX 23-APR-1992;  
PF 91GB-0008652.  
XX 23-APR-1991;  
PR 91GB-0008652.  
XX (ANTI-) ANTISOMA LTD.  
XX PA  
XX PI Courtenay-Luck NS;  
XX WPI, 1992-382045/46.  
DR N-PSDB; AAQ30197.  
XX  
XX New peptide EPPT (Glu-Pro-Pro-Thr) - selectively binds mucin  
PT expressed by epithelial tumours, used for guiding toxins or  
PT labels to tumours expressing mucin  
XX  
XX Example; Table 4, Page 30; 50pp; English.  
XX  
XX The antibody NM-2 is a murine monoclonal antibody class IgG.1,  
CC Lambda light chain, which has specificity for the mucin mol.  
CC The antibody reacts with about 95% of epithelial tumours and  
CC cross-reacts with normal mucin. NM-2 VH and VL genes were  
CC isolated using primers AAQ30193-Q30196. Both the heavy and the  
CC light chain genes for NM-2 were sequenced and the CDR sequences  
CC identified (see AAQ30197, AAQ30198).  
XX  
XX Sequence 107 AA;

Query Match 100.0%; Score 29; DB 13; Length 107;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYGVH 5  
| | | | |  
DB 28 sygvh 32

RESULT 5  
ID AAW31648 standard; Protein; 112 AA.  
XX  
XX AAW31648;  
XX  
XX 21-MAY-1998 (first entry)  
XX  
XX Monoclonal antibody CP.B8 heavy chain variable region.  
XX  
XX Cytokine receptor; gamma common chain; gc chain; human;  
KM blocking agent; monoclonal antibody; CP.B8; immunological disease;  
KM myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;  
KM insulin-dependent diabetes; inflammatory bowel disease;  
KM sympathetic ophthalmia; uveitis; allergy; asthma; infection;  
KM graft versus host disease; psoriasis; immunosuppressive; therapy;  
KM complementarity determining region; CDR.  
XX  
XX Mus musculus.

XX Key Location/Qualifiers  
XX Region 28..32  
FT /note= "CDR1"  
FT 47..61  
FT /note= "CDR2"  
FT 95..104  
FT /note= "CDR3"  
XX  
XX WO9743416-A1.  
XX  
XX 20-NOV-1997.  
XX  
XX 09-MAY-1997; 97WO-US07870.

XX 10-MAY-1996; 96US-0017466.  
PR (BIOJ ) BIOGEN INC.  
XX  
XX Benjamin CD, Burkiy LC, Hession C, Whitty A;  
XX WPI: 1998-008885/01.  
DR N-PSDB; AAT97441.  
XX  
XX Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
XX Claim 23; Page 81-82; 111pp; English.  
XX  
XX This polypeptide comprises the heavy chain variable region (VH) of  
CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma  
CC deposited at ATCC 12107, and which is specific for the gamma  
CC constant (gc) chain (see AAW31646) of human cytokine receptors. The  
CC invention provides compositions and methods for inhibiting cytokine  
CC signalling using gc chain blocking agents for the treatment of  
CC immunological diseases such as myasthenia gravis, rheumatoid  
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
CC allergy, asthma, parasitic infection, graft vs. host disease or  
CC psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab  
CC fragment and an antibody having a light chain variable region  
CC region CDR selected from those of CP.B8 VH or a heavy chain variable  
CC region CDR selected from those of CP.B8 VL (see AAW31647).  
XX  
XX Sequence 112 AA;

Query Match 100.0%; Score 29; DB 19; Length 112;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYGVH 5  
| | | | |  
DB 28 sygvh 32

RESULT 6  
ID AAR21268 standard; Protein; 113 AA.  
XX  
XX AAR21268;  
XX  
XX 21-MAY-1992 (first entry)  
XX  
XX Murine VH group 2 chain E specific for phox.  
DE  
XX Fd; bacteriophage; gene III; filamentous; phagemid; capsid; coat;  
KM plus; gp; binding; adsorption; gene VIII; diverse repertoire;  
KM specific binding pairs; replicable genetic display package.  
XX  
XX Synthetic.

XX Key Location/Qualifiers  
XX Binding-site 31..35  
FT /label= CDR1  
FT 50..65  
FT /label= CDR2  
FT 98..102  
FT /label= CDR3  
FT /note= "D/N-X-G-X-X motif "  
XX  
XX WO9201047-A.  
XX  
XX 23-JAN-1992.  
XX  
XX 10-JUL-1991; 91WO-GB01134.

XX 15-MAY-1991; 91GB-0010549.  
 PR 10-JUL-1990; 90GB-0015198.  
 PR 19-OCT-1990; 90GB-0022845.  
 PR 12-NOV-1990; 90GB-0024503.  
 PR 06-MAR-1991; 91GB-0004744.  
 XX  
 PA (CAMP-) CAMBRIDGE ANTIBODY.  
 PA (MEDI-) MED RES COUNCIL.  
 PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD,  
 PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;  
 PI Winter GP, Bonnett TP;  
 XX  
 DR WPI; 1992-056862/07.  
 XX  
 PT Producing members of specific binding pairs - by expression in  
 PT recombinant host cells with a secreting replicable genetic  
 PT display package.  
 XX  
 PS Example 21; Fig 24; 209pp; English.  
 XX  
 CC The VH sequence is one of eight (AAR21264-71) found to be expressed  
 CC from a single chain Fv library from an immunised mouse. The  
 CC library produces a diverse repertoire of antibody fragments specific  
 CC for 2-phenyl-5-oxazolone (phox). It was prepd. using cDNA generated  
 CC from mRNA from mice immunised with phox coupled to chicked serum  
 CC albumin. The VH and VL kappa sequences were separately amplified by  
 CC PCR (see AAQ23474-84) and ligated into fdCAT2 (see AAQ23463) for  
 CC expression on the phage surface as fusions with gene III. The  
 CC resulting library of clones was diverse. Twenty three hapten binding  
 CC clones were sequenced revealing the eight different VH genes (A-H)  
 CC in a variety of pairings with seven different VK genes (a-g) (see  
 CC AAR2164-92). Nearly all the VH genes belonged to gp 1, with only  
 CC one, "E", being of gp 2 (VHox1). Of the twenty three clones  
 CC sequenced, only one was of type "E". Most of the clones were VK-d  
 CC combinations. The kd of VH-B/VK-d for phox-GABA was 10 nM. Only two  
 CC other combinations (of eleven tested) were found to have higher  
 CC values. This suggests that phage bearing scfv fragments having  
 CC weak affinities can be selected with antigen, probably due to the  
 CC avidity of the multiple antibody heads on the phage.  
 CC See also AAR21260-307, 309-311; AAR22450, 565-581.  
 CC  
 XX  
 SO Sequence 113 AA;  
 Query Match 100.0%; Score 29; DB 13; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SYGVH 5  
 |||||  
 Db 31 sygvh 35  
 RESULT 7  
 AAW04595  
 ID AAW04595 standard; Protein; 115 AA.  
 XX  
 AC AAW04595;  
 XX  
 DT 12-AUG-1997 (first entry)  
 XX  
 DE Anti-DNA antibody 11f8 heavy chain variable region.  
 XX  
 KW Heavy chain; variable region; anti-DNA; monoclonal; antibody;  
 KW Mab 11f8; hairpin; diagnosis; inflammatory glomerulonephritis;  
 KW systemic lupus erythematosus; screening; treatment; prevention;  
 KW SLE; disease.  
 XX  
 OS Mus spp.  
 XX  
 FH Key Location/Qualifiers

FT Region 1..22  
 FT /label= framework\_I  
 FT 23..27  
 FT Region  
 FT /label= CDR\_I  
 FT 28..41  
 FT Region  
 FT /label= framework\_II  
 FT 42..58  
 FT Region  
 FT /label= CDR\_II  
 FT 59..90  
 FT Region  
 FT /label= framework\_III  
 FT 91..102  
 FT Region  
 FT /label= CDR\_III  
 FT 103..115  
 FT Region  
 FT /label= J\_region  
 XX  
 PN W09636361-A1.  
 XX  
 PD 21-NOV-1996.  
 XX  
 PE 16-MAY-1996; 96WO-US07113.  
 XX  
 PR 18-MAY-1995; 95US-0443540.  
 XX  
 PA (UNMI ) UNIV MICHIGAN.  
 XX  
 PI Glick GD, Swanson PC;  
 PI  
 XX  
 DR WPI; 1997-011854/01.  
 DR N-PSDB; AAT43740.  
 XX  
 PT Anti-DNA antibody which specifically binds DNA hairpin - useful to  
 PT develop prods. for diagnosis and treatment of disorders, e.g.  
 PT glomerulonephritis or systemic lupus erythematosus  
 XX  
 PS Example; Fig 7; 102pp; English.  
 XX  
 CC The present sequence is the heavy chain variable region of the  
 CC anti-DNA monoclonal antibody (Mab) 11f8, which has a high affinity  
 CC for single stranded DNA, low or no affinity for double stranded DNA  
 CC and specifically binds a DNA hairpin. The Mab can be used to diagnose  
 CC disorders associated with the pathological complexation of DNA,  
 CC e.g. inflammatory glomerulonephritis and systemic lupus  
 CC erythematosus. It can also be used to generate reagents to screen  
 CC for pharmaceutical agents, and treat and/or prevent an above  
 CC disorder.  
 CC Calt thymus DNA was used to immunise a MRJ-1pr mouse, spleen cells  
 CC from which were then fused with Sp2/0 myeloma cells to give  
 CC hybridomas producing the anti-DNA Mab. 11f8 was found to react  
 CC strongly with single stranded DNA and poly(dT).  
 CC  
 XX  
 SO Sequence 115 AA;  
 Query Match 100.0%; Score 29; DB 18; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SYGVH 5  
 |||||  
 Db 23 sygvh 27  
 RESULT 8  
 AAR07322  
 ID AAR07322 standard; protein; 116 AA.  
 XX  
 AC AAR07322;  
 XX  
 DT 22-JAN-1991 (first entry)  
 XX  
 DE VH domain of antibody D against tumour-associated antigens.  
 XX  
 KW Tumour-associated antigen; murine monoclonal antibody D;

KM vibrio cholerae neuraminidase-sensitive epitope; glioma;  
 KM meningioma; neurilemmoma; neuroblastoma; ganglioblastoma;  
 KM ganglioneuroma; diagnosis.  
 XX  
 OS Mus musculus.  
 XX  
 PN EP388914-A.  
 XX  
 PD 26-SEP-1990.  
 XX  
 PF 21-MAR-1990; 90EP-0105322.  
 XX  
 PR 24-MAR-1989; 89DE-3909799.  
 XX  
 PA (BEHW ) BEHRINGERK AG.  
 XX  
 PI Bosslet K, Seemann G, Sedlacek HH;  
 XX  
 DR WPI: 1990-291873/39.  
 DR N-PSDB; AAQ06229.  
 XX  
 PT Monoclonal antibodies to tumour associated antigens - used for  
 PT diagnosis of malignant tumours etc.  
 PS Disclosure; Page 14; 18pp; German.  
 XX  
 CC Antibody D is produced as described in EP-141079 and binds to  
 CC a vibrio cholerae neuraminidase-sensitive epitope on ganglioside GD2  
 CC which occurs in gliomas, meningiomas, neurilemmomas,  
 CC neuroblastomas, ganglioblastomas and ganglioneuromas.  
 CC They are useful in tumour diagnosis and therapy.  
 CC See also AAQ06230 for VK of Mab D, AAQ07312-15 for Mab A and B and  
 CC AAQ06227-28 for Mab C.  
 XX  
 SQ Sequence 116 AA;  
 OY  
 Db 1 SYGVH 5  
 11111  
 26 sygvh 30  
 RESULT 9  
 AAR28287  
 ID AAR28287 standard; Protein: 116 AA.  
 XX  
 AC AAR28287;  
 XX  
 DT 04-APR-1993 (first entry)  
 XX  
 DE Sequence of the VH region of the Kabat human heavy chain  
 DE subgroup II of anti-mucin monoclonal antibody.  
 XX  
 KM Antibody; PCR primer; variable heavy chain.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT 26..30 Location/Qualifiers  
 FT /label= CDR 1  
 FT 44..60  
 FT /label= CDR 2  
 FT 93..101  
 FT Region /label= CDR 3  
 XX  
 PN WO9218534-A.  
 XX  
 PD 29-OCT-1992.  
 XX

PF 23-APR-1992; 92WO-GB00746.  
 XX  
 PR 23-APR-1991; 91GB-0008652.  
 XX  
 PA (ANTI-) ANTISOMA LTD.  
 XX  
 PI Courtenay-luck NS;  
 XX  
 DR WPI: 1992-382045/46.  
 DR N-PSDB; AAQ30192.  
 XX  
 PT New peptide EPPt (Glu-Pro-Pro-Thr) - selectively binds mycin  
 PT expressed by epithelial tumours, used for guiding toxins or  
 PT labels to tumours expressing mucin  
 XX  
 PS Example; Table 2, Page 28; 50pp; English.  
 XX  
 CC Clone B is a lymphoblastoid cell line (secreting antibody directed  
 CC against a tumour-associated mucin mol.) derived from the EBV-  
 CC transforming and cloning of a patient's peripheral blood B-cells.  
 CC After DNA isolation, the polymerase chain reaction (PCR) was  
 CC employed, using oligonucleotide primers specific for the variable  
 CC light and heavy chains of immunoglobulins (see AAQ30065, AAQ30066).  
 CC From this mini-prep the VH gene of the human antibody, which is  
 CC designated as clone-B was isolated. Sequencing of the gene encoding the  
 CC VH region of clone B provided a sequence consistent with the Kabat  
 CC human heavy chain subgroup II of anti-mucin monoclonal antibody.  
 XX  
 SQ Sequence 116 AA;  
 OY  
 Db 1 SYGVH 5  
 11111  
 26 sygvh 30  
 RESULT 10  
 AAW14490  
 ID AAW14490 standard; Protein: 116 AA.  
 XX  
 AC AAW14490;  
 XX  
 DT 28-JAN-1997 (first entry)  
 XX  
 DE Monoclonal antibody D VH.  
 XX  
 KM heavy; light chain; monoclonal antibody; antigen 3; marker; melanoma;  
 KM permanent human tumour cell line; tumour-associated antigen; epitope;  
 KM gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;  
 KM antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.  
 XX  
 OS Synthetic.  
 XX  
 PN EP727436-A1.  
 PN  
 PD 21-AUG-1996.  
 XX  
 PF 21-MAR-1990; 90EP-0105322.  
 XX  
 PR 24-MAR-1989; 89DE-3909799.  
 XX  
 PA (BEHW ) BEHRINGERK AG.  
 XX  
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 XX  
 DR WPI: 1996-372836/38.  
 DR N-PSDB; AAT63507.  
 XX  
 PT Monoclonal antibody to tumour-associated antigen - useful as

PT gastrointestinal tumour marker  
 XX  
 PS Disclosure; Page 14; 19pp; German.  
 XX  
 CC AAM14490-91 are the heavy and light chains (respectively) of monoclonal  
 CC antibody (Mab) D. Mab D recognises Vibrio cholera  
 CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma  
 CC cell line. Mabs A, B and C (see AAM14484-89) are mentioned in the  
 CC specification, but are not part of the claims. Mabs A and B recognise  
 CC antigens 3 and 11 resp., of a permanent human tumour cell line. Mab C  
 CC also recognises an epitope of a tumour-associated antigen. These antigens  
 CC occur at high concns. in the serum of patients with gastrointestinal  
 CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers  
 CC for diagnostic or therapeutic purposes.  
 CC  
 XX  
 SQ Sequence 116 AA;  
 OY 1 SYGVH 5  
 DB 26 sygvh 30  
 Query Match 100.0%; Score 29; DB 17; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SYGVH 5  
 DB 26 sygvh 30  
 RESULT 11  
 AAR9877  
 ID AAR9877 standard; Protein: 116 AA.  
 AC AAR9877;  
 XX 28-JAN-1997 (first entry)  
 DT  
 XX Monoclonal antibody D VH.  
 DE  
 XX Monoclonal antibody; Mab; epitope: tumour-associated antigen;  
 KW marker: antigen.  
 XX  
 XX Synthetic.  
 OS  
 XX EP727435-A1.  
 PN  
 XX 21-AUG-1996.  
 PD  
 XX 21-MAR-1990; 90EP-0105322.  
 PF  
 XX 24-MAR-1989; 89DE-3909799.  
 PR  
 PA (BEHW ) BEHRINGWERKE AG.  
 XX  
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 XX  
 XX WPI: 1996-372835/38.  
 DR N-PSDB; AAT3665.  
 DR  
 XX Monoclonal antibody to tumour-associated antigen - useful as  
 PT gastrointestinal tumour marker  
 XX  
 PS Disclosure; Page 14; 19pp; German.  
 XX  
 CC Mab C (AAT36659-T36660) is a monoclonal antibody that recognises an  
 CC epitope of a tumour-associated antigen occurring at high concn. in  
 CC the serum of patients with gastrointestinal tumours, e.g. pancreatic  
 CC carcinoma, and is thus useful as a tumour marker for diagnostic or  
 CC therapeutic purposes.  
 CC Mabs A, B and D are mentioned in the specification, but are not  
 CC part of the claims.  
 CC Mab A (AAT36661-T36662) recognises antigen 3 of permanent human  
 CC tumour cell line.  
 CC Mab B (AAT36663-T36664) recognises antigen 11 of permanent human  
 CC tumour cell line.

CC Mab D (AAT36665-T36666) recognises a Vibrio cholera neuraminidase-  
 CC resistant epitope of ganglioside GD2, from a human melanoma cell  
 CC line.  
 CC  
 XX  
 SQ Sequence 116 AA;  
 OY 1 SYGVH 5  
 DB 26 sygvh 30  
 Query Match 100.0%; Score 29; DB 17; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SYGVH 5  
 DB 26 sygvh 30  
 RESULT 12  
 AAR25728  
 ID AAR25728 standard; Protein: 119 AA.  
 AC AAR25728;  
 XX 13-JAN-1993 (first entry)  
 DT  
 XX Humanised VH region of the mouse mik-beta1 antibody.  
 DE  
 XX Murine: immunoglobulin; CDR; non immunogenic; interleukin-2;  
 KW IL-2; heavy chain; variable region; framework; p75; human; lay.  
 XX  
 XX Mus musculus.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 31..35  
 FH Region /note= "CDR"  
 FT 50..65  
 FT Region /note= "CDR"  
 FT 98..108  
 FT Region /note= "CDR"  
 FT  
 FT Misc-difference 1 /note= "mutated residue"  
 FT 29 /note= "mutated residue"  
 FT Misc-difference 30 /note= "mutated residue"  
 FT 49 /note= "mutated residue"  
 FT Misc-difference 72 /note= "mutated residue"  
 FT 73 /note= "mutated residue"  
 FT Misc-difference 84 /note= "mutated residue"  
 FT 89 /note= "mutated residue"  
 FT Misc-difference 90 /note= "mutated residue"  
 FT  
 FT W09211018-A.  
 PN  
 XX  
 PD 09-JUL-1992.  
 XX  
 XX 19-DEC-1991; 91WO-US09711.  
 PF  
 XX 19-DEC-1990; 90US-0634278.  
 PR  
 XX (PROT-) PROTEIN DESIGN LABS INC.  
 PA  
 XX Co MS, Coelinhg KL, Landolfi NF, Queen CL, Schneider WP;  
 PI  
 XX WPI: 1992-249842/30.  
 DR  
 XX New immunoglobulin(s) having murine CDRs in human framework  
 PT regions - have lower antigenicity; useful for treating e.g. HSV,

CC The invention provides a monoclonal antibody (Mab) designated 1A7, which  
CC elicits an anti-GD2 (tumor-associated antigen) immunological response in  
CC humans. Mab 1A7 has defined light and heavy chain variable region  
CC sequences. The Mab 1A7 and polypeptides can be used for eliciting an

CC This sequence represents the heavy chain from the K5H8 antibody. The  
CC invention relates to a humanized antibody framework motif that is  
CC selected from a human library based upon comparison to a murine antibody,  
CC and the heavy chain is encoded by the VH gene of K5H8, and the light  
CC chain is encoded by the VL gene of Tr1.6. The humanized antibodies can be  
CC used in a variety of ways, and may be expressed intracellularly. For  
CC example, they can be directed against the activation domain of HIV-1, and  
CC so inhibit HIV-1 replication. They can also be used extracellularly to  
CC target Tat, and, when combined with a detectable label, be used to  
CC measure levels of Tat for diagnosis or prognosis, and to diagnose  
CC susceptibility to HIV infection. In order to minimize risk of evoking an  
CC immune response, murine anti-tat antibodies are humanized. However, prior  
CC art methods have resulted in humanized antibodies that are not as  
CC effective as the original antibody. Frequently, the humanized antibody  
CC will retain some murine amino acids. It would be desirable to have a  
CC framework motif that produces an antibody having a protective efficiency  
CC comparable to the murine antibody. This need is met by the humanized  
CC antibodies. By retaining no muting amino acids in the humanized

CC framework, a more effective antibody against HIV is produced, compared to  
CC prior art methods.  
XX  
SQ Sequence 119 AA;

OY 1 SYGVH 5  
Db 31 sygvh 35

Query Match 100.0%; Score 29; DB 20; Length 119;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: August 20, 2002, 13:15:09  
Job time: 103 sec

OY 1 SYGVH 5  
Db 31 sygvh 35

RESULT 15  
AAB69660  
ID AAB69660 standard; Protein: 119 AA.  
XX  
AC AAB69660;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Human lay antibody heavy chain SEQ ID NO: 37.  
XX  
KW Humanised immunoglobulin; mouse; human; antibody; heavy chain; diabetes;  
KW light chain; graft versus host disease; transplant; autoimmune disease;  
KW multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;  
KW myasthenia gravis; herpes infection; myeloid leukaemia; CMV infection.  
XX  
OS Homo sapiens.  
XX  
PN US6180370-B1.  
XX  
PD 30-JAN-2001.  
XX  
PE 07-JUN-1995; 95US-0484537.  
XX  
PR 28-DEC-1988; 88US-0290975.  
PR 13-FEB-1989; 89US-0310252.  
PR 28-SEP-1990; 90US-0590274.  
PR 19-DEC-1990; 90US-0634278.  
XX  
PA (PROT-) PROTEIN DESIGN LABS INC.  
XX  
PI Queen CL, Selick HE;  
XX  
DR WPI: 2001-190856/19.  
XX  
PT Producing humanized immunoglobulin, involves producing a cell  
PT containing DNA segments encoding humanized heavy and light chain  
PT variable regions, and expressing the DNA segments in the cell -  
XX  
PS Example 5; Fig 26; 145pp; English.  
XX  
CC The present invention describes a method of producing humanised  
CC immunoglobulins involving expressing in a cell a nucleic acid encoding a  
CC humanised version of an immunoglobulin. This is obtained by comparing a  
CC donor and human immunoglobulin and producing a combined antibody which  
CC contains part of each. These are useful in the treatment of  
CC graft-versus-host disease, transplant rejection, autoimmune diseases such  
CC as diabetes, rheumatoid arthritis, myasthenia gravis, multiple sclerosis  
CC and systemic lupus erythematosus, herpes infections, CMV virus infections  
CC and myeloid leukaemia. The present sequence is an antibody used to  
CC demonstrate the method of the invention.  
XX  
SQ Sequence 119 AA;

Query Match 100.0%; Score 29; DB 22; Length 119;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:26 ; Search time 46.28 Seconds  
(Without alignments)  
10.381 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_28\_32

Perfect score: 29

Sequence: 1 SYGVH 5

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	76	2 AH2120	hypothetical prote
2	29	100.0	90	2 A49042	Ig heavy chain V r
3	29	100.0	95	2 S17605	Ig heavy chain V r
4	29	100.0	97	2 S55372	Ig heavy chain V r
5	29	100.0	98	2 S21812	Ig heavy chain V r
6	29	100.0	100	2 A25913	Ig heavy chain pre
7	29	100.0	100	2 S14490	Ig heavy chain V r
8	29	100.0	101	2 S03466	Ig heavy chain V r
9	29	100.0	101	2 PH1021	Ig heavy chain V r
10	29	100.0	105	2 PH1020	Ig heavy chain V r
11	29	100.0	106	2 S26322	Ig heavy chain V r
12	29	100.0	106	2 S14489	Ig heavy chain V r
13	29	100.0	107	2 S14492	Ig heavy chain V r
14	29	100.0	107	2 S14491	Ig heavy chain V r
15	29	100.0	108	2 S07331	Ig heavy chain V r
16	29	100.0	109	2 PH1023	Ig heavy chain V r
17	29	100.0	109	2 PH1025	Ig heavy chain V r
18	29	100.0	111	2 PH1019	Ig heavy chain V r
19	29	100.0	112	2 PH1022	Ig heavy chain V r
20	29	100.0	112	2 S11098	Ig heavy chain V r
21	29	100.0	112	2 S11100	Ig heavy chain V r
22	29	100.0	112	2 S11108	Ig heavy chain V r
23	29	100.0	113	2 S11101	Ig heavy chain V r
24	29	100.0	114	2 S26321	Ig heavy chain V r
25	29	100.0	114	2 S11099	Ig heavy chain V r
26	29	100.0	114	2 S11104	Ig heavy chain V r
27	29	100.0	114	2 S11106	Ig heavy chain V r
28	29	100.0	114	2 S11105	Ig heavy chain V r
29	29	100.0	115	2 S26470	Ig heavy chain V r

30	29	100.0	115	2 S11107	Ig heavy chain V r
31	29	100.0	116	1 G1MS10	Ig heavy chain pre
32	29	100.0	116	2 A33932	Ig mu chain precu
33	29	100.0	116	2 S11102	Ig heavy chain V r
34	29	100.0	117	2 S10111	Ig heavy chain V r
35	29	100.0	127	2 B31807	Ig heavy chain V r
36	29	100.0	135	2 S31913	Ig gamma-2A chain
37	29	100.0	141	2 S52446	Ig heavy chain V r
38	29	100.0	144	2 S11244	Ig gamma-2a chain
39	29	100.0	240	2 D89916	dihydrodipicolinat
40	29	100.0	285	2 S29306	poly(3-hydroxylaika
41	29	100.0	288	2 E89903	hypothetical prote
42	29	100.0	305	2 T49720	hypothetical prote
43	29	100.0	397	2 F90182	hypothetical prote
44	29	100.0	587	2 AE0487	probable hemolysin
45	29	100.0	657	2 T12969	hypothetical prote

## ALIGNMENTS

RESULT 1  
AH2120  
hypothetical protein asl2519 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence-revision 14-Dec-2001 #text-change 11-Jan-2002  
C:Accession: AH2120  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iridu  
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2120  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-76 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA874218.1; PID:g17131611; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: asl2519

Query Match 100.0%; Score 29; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5  
|||||  
Db 31 SYGVH 35

RESULT 2  
A49042  
Ig heavy chain V region (anti-DNA) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 19-Dec-1993 #sequence-revision 18-Nov-1994 #text-change 23-Jul-1999  
C:Accession: A49042  
R:Taki, S.; Hirose, S.; Kinoshita, K.; Nishimura, H.; Shimamura, T.; Hamuro, J.; Shlr  
Eur. J. Immunol. 22, 987-992, 1992  
A:Title: Somatic mutation of anti-DNA antibody clonally related to germ-line enco  
A:Reference number: A49042; MUID:92201320  
A:Accession: A49042  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-90 <TAK>  
A:Cross-references: GB:S92270; NID:g247954; PIDN:AAB21904.1; PID:g247955  
A:Experimental source: NZB, NZW, liver  
C:Note: sequence extracted from NCBI backbone (NCBI:92270, NCBI:92271)  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:8-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 90;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYGVH 5  
|||||  
Db 24 SYGVH 28

RESULT 3  
S17605  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 06-Jun-1997  
C:Accession: S17605  
R:Clackson, T.; Hoogenboom, H.R.; Griffiths, A.D.; Winter, G.  
Nature 352, 624-628, 1991  
A:Title: Making antibody fragments using phage display libraries.  
A:Reference number: S17230; MUID:91326098  
A:Accession: S17605  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-95 <CLa>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 24 SYGVH 28

RESULT 4  
S53372  
Ig heavy chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Sep-1995 #sequence\_revision 19-Oct-1995 #text\_change 23-Jul-1999  
C:Accession: S53372  
R:Well, M.; Grimmer, K.; Simon, D.; Lefranc, G.; Pau, B.; Lefranc, M.P.  
submitted to the EMBL Data Library, November 1990  
A:Description: Sequence of the 4C1 antibody heavy and light chain variable regions.  
A:Reference number: S55366  
A:Accession: S53372  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-97 <MEI>  
A:Cross-references: EMBL:X56177; NID:9854282; PIDN:CAA39638.1; PID:9854283  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 23 SYGVH 27

RESULT 5  
S21812  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 16-Aug-1996  
C:Accession: S21812  
R:Mocklat, R.

submitted to the EMBL Data Library, March 1991  
A:Reference number: S21812  
A:Accession: S21812  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-98 <MOC>  
A:Cross-references: EMBL:X58634  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 98;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 23 SYGVH 27

RESULT 6  
A25913  
Ig heavy chain precursor V region (BFL2) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 21-Jan-2000  
C:Accession: A25913  
R:Lawler, A.M.; Lin, P.S.; Gearhart, P.J.  
Proc. Natl. Acad. Sci. U.S.A. 84, 2454-2458, 1987  
A:Title: Adult B-cell repertoire is biased toward two heavy-chain variable-region genes  
A:Reference number: A94148; MUID:87175692  
A:Accession: A25913  
A:Molecule type: DNA  
A:Residues: 1-100 <LA>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:19-100/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 35 SYGVH 39

RESULT 7  
S14490  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S14490  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL Data Library, March 1991  
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in V  
A:Reference number: S14484  
A:Accession: S14490  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-100 <CH>  
A:Cross-references: EMBL:X58647; NID:951283; PIDN:CAA41504.1; PID:951284  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5



Db 31 SYGVH 35

## RESULT 8

S03466  
Ig heavy chain V region (Vcb) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 16-Aug-1996  
C:Accession: S03466  
R:Rech, M.G.; Jackson, S.; Alt, F.W.  
EMBO J. 5, 2131-2138, 1986  
A:Title: V(H)DJ(H) formation and DJ(H) replacement during pre-B differentiation: non-rat  
A:Reference number: S03466; MUID:87053850  
A:Accession: S03466  
A:Molecule type: DNA  
A:Residues: 1-101 <RET>  
A:Cross-references: EMBL:X04348  
A>Note: the authors translated the codon ATC for residue 113 as Ser  
A>Note: this sequence was determined from the nonfunctional differentiated gene  
C:Genetics:  
A:Map position: 12  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:18-100/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYGVH 5  
Db 34 SYGVH 38

## RESULT 9

PH1021  
Ig heavy chain V region (clone 165.52) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C:Accession: PH1021  
R:Fillman, D.M.; Jou, N.T.; Hall, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992  
A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective B  
A:Reference number: PH0971; MUID:92381444  
A:Accession: PH1021  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-101 <TIL>  
A:Experimental source: B cell, strain (NZB x NZW)F1  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:7-89/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYGVH 5  
Db 23 SYGVH 27

## RESULT 10

PH1020  
Ig heavy chain V region (clone 165.41) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C:Accession: PH1020  
R:Fillman, D.M.; Jou, N.T.; Hall, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992

A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective  
A:Reference number: PH0971; MUID:92381444

A:Accession: PH1020  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-105 <TIL>  
A:Experimental source: B cell, strain (NZB x NZW)F1  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:14-95/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 105;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYGVH 5  
Db 29 SYGVH 33

## RESULT 11

S26322  
Ig heavy chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Mar-1998 #sequence\_revision 19-Mar-1998 #text\_change 21-Jan-2000  
C:Accession: S26322  
R:Stark, S.E.; Caton, A.J.  
J. Exp. Med. 174, 613-624, 1991  
A:Title: Antibodies that are specific for a single amino acid interchange in a proteol  
A:Reference number: S26309; MUID:91341421  
A:Accession: S26322  
A:Molecule type: mRNA  
A:Residues: 1-106 <STA>  
A:Cross-references: EMBL:X59182  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:9-91/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYGVH 5  
Db 25 SYGVH 29

## RESULT 12

S14489  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S14489  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL data library, March 1991  
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in V  
A:Reference number: S14484  
A:Accession: S14489  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-106 <CHE>  
A:Cross-references: EMBL:X58646; NID:951281; PIDN:CAA41503.1; PID:951282  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 31 SYGVH 35

## RESULT 13

SI4492  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: SI4492  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL Data Library, March 1991  
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in VH C  
A:Reference number: SI4484  
A:Accession: SI4492  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-107 <CHE>  
A:Cross-references: EMBL:X58649; NID:g51287; PIDN:CAA41506.1; PID:g51288  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 31 SYGVH 35

## RESULT 14

SI4491  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: SI4491  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL Data Library, March 1991  
A:Description: Natural polyclonal antibodies differ from Ag-induced antibodies in VH C  
A:Reference number: SI4484  
A:Accession: SI4491  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-107 <CHE>  
A:Cross-references: EMBL:X58648; NID:g51285; PIDN:CAA41505.1; PID:g51286  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 31 SYGVH 35

## RESULT 15

S07331  
Ig heavy chain V region (clone NO2-12.4) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Feb-1997 #sequence\_revision 18-Jul-1997 #text\_change 21-Jan-2000  
C:Accession: S07331  
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.  
Nature 304, 320-324, 1983  
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone  
A:Reference number: S07331; MUID:83271467

A:Accession: S07331  
A:Molecule type: mRNA  
A:Residues: 1-108 <KA>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 29; DB 2; Length 108;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 31 SYGVH 35

Search completed: August 20, 2002, 13:16:02  
Job time: 156 sec

2002-08-20 13:05:56

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:31 ; Search time 21.48 Seconds  
(without alignments)  
9.013 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_28\_32  
Perfect score: 29  
Sequence: 1 SYGVH 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	116	1 HV45_MOUSE	P01821 mus musculus
2	29	100.0	230	1 DAPB_STRAU	Q92111 staphylococ
3	29	100.0	1856	1 MGA_HUMAN	Q43451 homo sapien
4	29	96.6	187	1 PAAD_HELPJ	Q92161 helicobacte
5	28	96.6	187	1 PAAD_HELPJ	Q26011 helicobacte
6	28	96.6	840	1 MUTS_LACIA	Q92499 lactococcus
7	28	96.6	844	1 HEXA_STRPN	P01564 streptococc
8	28	96.6	851	1 MUTS_STRPN	Q99X18 streptococc
9	28	96.6	858	1 MUTS_BACSU	P49849 bacillus su
10	26	89.7	89	1 RS15_MYCLE	Q33267 mycobacteri
11	26	89.7	139	1 Y965_MYCTU	P11734 locusta mig
12	26	89.7	148	1 CU08_LOCM1	Q15140 mycobacteri
13	26	89.7	185	1 3MGR_RHIME	Q92111 ribzobium m
14	26	89.7	185	1 3MGR_RHIME	Q92111 ribzobium m
15	26	89.7	195	1 GRP3_DAUCA	P37705 daucus caro
16	26	89.7	198	1 YN00_YEAST	P33846 saccharomyc
17	26	89.7	211	1 THIE_ECOLI	P30137 escherichia
18	26	89.7	211	1 THIE_SALTY	Q91918 salmonella
19	26	89.7	412	1 PT56_YEAST	P52270 saccharomyc
20	26	89.7	435	1 YFS3_CAEEL	P49191 caenorhabdi
21	26	89.7	452	1 HOS2_YEAST	P33096 saccharomyc
22	26	89.7	474	1 PNTB_HABIN	P33010 haemophilus
23	26	89.7	524	1 SYK_CERSY	Q74059 cenarchoaem
24	26	89.7	534	1 NU2M_NEUCR	Q35140 neurospora
25	26	89.7	658	1 CTR2_HUMAN	P52569 homo sapien
26	26	89.7	676	1 TKT7_CRAPL	Q42677 craterostilg
27	26	89.7	679	1 TKTA_CRAPL	Q42675 craterostilg
28	26	89.7	704	1 G1GB_YEAST	P32775 saccharomyc
29	26	89.7	736	1 ALD_MOUSE	P48410 mus musculu
30	26	89.7	795	1 SYFB_VIRCH	Q93666 mus musculu
31	26	89.7	803	1 RIR1_CRYPV	Q61065 cryptospori
32	26	89.7	804	1 RIR1_PLAFG	P50647 plasmodium
33	26	89.7	806	1 RIR1_PLAF4	P50648 plasmodium

34	26	89.7	835	1 UREA_SCHPO	O00084 schizosacch
35	26	89.7	891	1 MUTS_RICPR	Q924m1 rickettsia
36	26	89.7	1051	1 YC94_HUMAN	Q924m2 homo sapien
37	26	89.7	1062	1 YAL3_HUMAN	Q924m3 homo sapien
38	26	89.7	1465	1 ATW2_HUMAN	P54296 homo sapien
39	26	89.7	1593	1 TYCC_BACBR	P58397 homo sapien
40	26	89.7	6486	1 TYCC_BACBR	Q27225 carclanus ma
41	25	86.2	113	1 MTH_CARRA	Q27225 carclanus ma
42	25	86.2	142	1 YN8P_YEAST	P53737 saccharomyc
43	25	86.2	167	1 YBL7_SCHPO	Q924m4 schizosacch
44	25	86.2	208	1 VEP_NPVPD	O10318 oryza pseu
45	25	86.2	220	1 YADF_ECOLI	P36857 escherichia

## ALIGNMENTS

RESULT	ID	HV45_MOUSE	STANDARD	PRT	116 AA.
1	HV45_MOUSE				
1	P01821				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	Ig heavy chain V region MC101 precursor.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=82075900; PubMed=6273429;				
RA	Kataoka T., Nikaido T., Miyata T., Moriwa K., Honjo T.;				
RT	"The nucleotide sequences of rearranged and germline immunoglobulin				
RT	VH genes of a mouse myeloma MC101 and evolution of VH genes in				
RT	mouse."				
RL	J. Biol. Chem. 257:277-285(1982).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; J00502; AAA38515.1; -				
DR	PIR; A02096; GIMS10.				
DR	InterPro; IPR003006; Ig_MHC.				
DR	InterPro; IPR003596; Ig_V.				
DR	Pfam; PF00047; Ig; 1.				
DR	SMART; SM00406; IGV; 1.				
KW	Immunoglobulin V region; Signal.				
FT	SIGNAL				
FT	1				
FT	20				
FT	116				
FT	NON_TER				
SO	SEQUENCE				
	116 AA; 12593 MW; 8079A6EE7C552B3E CRC64;				
Query Match					
Best Local Similarity 100.0%; Score 29; DB 1; Length 116;					
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	1 SYGVH 5				
DB	50 SYGVH 54				
RESULT 2					
ID	DAPB_STAU	STANDARD:	PRT:	230 AA.	
AC	Q9E211;				
DT	16-OCT-2001 (Rel. 40, Created)				

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dihydropyridicolinate reductase (EC 1.3.1.26) (DHPR).
GN DABP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8325-4;
RA Wiltshire M.D., Burnham M.K., Foster S.J.;
RT "Identification and analysis of Staphylococcus aureus components
RT involved during growth in sera.";
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydropyridicolinate + NAD(P)(+) =
CC 2,3-dihydropyridicolinate + NAD(P)H.
CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC semialdehyde; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DIHYDROPYRIDICOLINATE REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF306669; AAC42247.1; -.
CC InterPro: IPR000846; DABP.
CC DR Pfam: PF01113; DABP. 1.
CC PRODOM: PD004105; DABP. 1.
CC PROSITE: PS01298; DABP; FALSE NEG.
KW Diaminopimelate biosynthesis; lysine biosynthesis; Oxidoreductase;
KW MAP.
SQ SEQUENCE 230 AA; 25523 MW; 10FCF79D7AF84A93 CRC64;
OY 1 SYGVH 5
Db 97 SYGVH 101
Query Match 100.0%; Score 29; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RN [2]
RX CHARACTERIZATION.
RP TISSUE=Small intestine mucosa;
RC MEDLINE=89066802; PubMed=3143729;
RX Naim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase.";
RL J. Biol. Chem. 263:19709-19717(1988).
RN [3]
RP SULFATION.
RX MEDLINE=88082658; PubMed=3121301;
RA Dantelisen E.M.;
RT "Tyrosine sulfation, a post-translational modification of microvillar
RT enzymes in the small intestinal enterocyte.";
RL EMBJ J. 6:2891-2896(1987).
CC -1- FUNCTION: MAY SERVE AS AN ALTERNATE PATHWAY FOR STARCH DIGESTION
CC WHEN LUMINAL ALPHA-AMYLASE ACTIVITY IS REDUCED BECAUSE OF
CC IMMATUREITY OR MALNUTRITION. MAY PLAY A UNIQUE ROLE IN THE
CC DIGESTION OF MALTED DIETARY OLIGOSACCHARIDES USED IN FOOD
CC MANUFACTURING.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. BRUSH BORDER.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN SMALL INTESTINE, GRANULOCYTE, AND
CC KIDNEY BUT NOT IN SALIVARY GLAND OR PANCREAS.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- PTM: DOES NOT UNDERGO INTRACELLULAR OR EXTRACELLULAR PROTEOLYTIC
CC CLEAVAGE.
CC -1- PTM: SULFATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 2 P-TYPE (TREFOLI) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF016833; AAC39568.1; -.
CC MIM: 154360; -.
CC DR InterPro: IPR000322; Glyco_hydro_31.
CC DR InterPro: IPR000519; P_trefoli.
CC DR Pfam: PF01055; Glyco_hydro_31; 2.
CC DR Pfam: PF00088; trefoli; 2.
CC DR SMART: SM00018; P; 2.
CC DR PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 2.
CC DR PROSITE: PS00025; P_TREFOLI; 1.
CC KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;
CC Glycosidase; Repeat; Signal-anchor; Sulfation.
CC INIT_MET 0
CC TRANSMEM 1
CC DOMAIN 13 12
CC FT DOMAIN 34 1856 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 37 83 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT DOMAIN 88 132 (POTENTIAL).
CC FT DOMAIN 197 914 LUMENAL (POTENTIAL).
CC FT DOMAIN 952 998 SER/THR-RICH.
CC FT DOMAIN 1066 1812 P-TYPE 1.
CC FT ACT_SITE 528 528 MALTAse.
CC FT ACT_SITE 1419 1419 GLUCOAMYLASE.
CC FT DISULFID 89 117 BY SIMILARITY.
CC FT DISULFID 100 116 BY SIMILARITY.
CC FT DISULFID 111 129 BY SIMILARITY.
CC FT DISULFID 965 982 BY SIMILARITY.
CC FT DISULFID 977 995 BY SIMILARITY.
CC FT MOD_RES 415 415 SULFATION (POTENTIAL).

```

FT MOD\_RES 424 424 Sulfation (POTENTIAL).  
 FT MOD\_RES 1281 1281 Sulfation (POTENTIAL).  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 826 826 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 884 884 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 976 976 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1254 1254 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1363 1363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1387 1387 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1602 1602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1671 1671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1841 1841 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1846 1846 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 1856 AA; 209571 MM; 4A1903B82E768004 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 1856;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 |||||  
 DB 1165 SYGVH 1169

RESULT 4  
 ID PAAD\_HELPY STANDARD; PRT; 187 AA.  
 AC 09ZJEB;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).  
 GN JHP1369.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
 RA Tummimo P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 RT gastric pathogen Helicobacter pylori.";  
 RL Nature 397:176-180(1999).  
 CC -1- SIMILARITY: BELONGS TO THE POLYPRENTL P-HYDROXYBENZONATE /  
 CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AEO01560; AAD06958.1; -  
 CC InterPro: IPR003382; Flavoprotein.  
 DR Pfam: PF02441; Flavoprotein; 1.  
 DR Hypothetical protein; Lyase; Decarboxylase; Complete proteome.  
 KW

SO SEQUENCE 187 AA; 20598 MM; D9BFAD9029FA396A CRC64;

Query Match 96.6%; Score 28; DB 1; Length 187;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 |||||  
 DB 76 SYGVH 80

RESULT 5  
 ID PAAD\_HELPY STANDARD; PRT; 187 AA.  
 AC 026011;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).  
 GN HPI476.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kervavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalas H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori.";  
 RL Nature 388:539-547(1997).  
 CC -1- SIMILARITY: BELONGS TO THE POLYPRENTL P-HYDROXYBENZONATE /  
 CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AEO00647; AAD08515.1; -  
 CC TrIR; HPI476; -  
 DR InterPro: IPR003382; Flavoprotein.  
 DR Pfam: PF02441; Flavoprotein; 1.  
 KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.

SO SEQUENCE 187 AA; 20586 MM; 92E1361FDA49AF80 CRC64;

Query Match 96.6%; Score 28; DB 1; Length 187;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 |||||  
 DB 76 SYGVH 80

RESULT 6  
 ID MUTS\_LACLA STANDARD; PRT; 840 AA.  
 AC 09CDK9;

```
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein muts.
GN MUTS OR HEXA OR IL2210.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
CC Lactococcus.
CC NCBI_TaxID=1360;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE006450; AAK06308.1; -
DR InterPro; IPR000432; MUTS_C.
DR InterPro; IPR002863; MUTS_N.
DR Pfam; PF00488; MUTS_C; 1.
DR Pfam; PF01624; MUTS_N; 1.
DR ProDom; PD001263; MUTS_C; 1.
DR SMART; SM00533; MUTSd; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
DR DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT NP_BIND 601 608 ATP (POTENTIAL).
SQ SEQUENCE 840 AA; 94272 MW; 867EFE4F82616DC2 CRC64;

OY 1 SYGVH 5
DB 757 SYGII 761

Query Match 96.6%; Score 28; DB 1; Length 840;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
RT Streptococcus pneumoniae and homologs of hexa to muts of Escherichia
RT coli and Salmonella typhimurium";
RT J. Bacteriol. 170:190-196(1988).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson M.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Kouri H., Wolf A.M., Ueberback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step.
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 759.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M18729; AAA88597.1; -
DR EMBL; AE007496; -; NOT_ANNOTATED_CDS.
DR PIR; C28667; C28667.
DR TIGR; SP2076; -
DR InterPro; IPR000432; MUTS_C.
DR InterPro; IPR002863; MUTS_N.
DR Pfam; PF00488; MUTS_C; 1.
DR Pfam; PF01624; MUTS_N; 1.
DR ProDom; PD001263; MUTS_C; 1.
DR SMART; SM00533; MUTSd; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
DR DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT NP_BIND 602 609 ATP (POTENTIAL).
SQ SEQUENCE 844 AA; 94848 MW; 59C1B963D90B6EE2 CRC64;

OY 1 SYGVH 5
DB 758 SYGII 762

Query Match 96.6%; Score 28; DB 1; Length 844;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```



RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC: 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Seate S., Svorov A.N., Kenton S., Lal H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
CC -1- FUNCTION: This protein is involved in the repair of mismatches in  
CC DNA. It is possible that it carries out the mismatch recognition  
CC step. This protein has a weak ATPase activity (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MTS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE006633; AAK4785.1; -  
DR InterPro: IPR000432; Muts\_C.  
DR InterPro: IPR002863; Muts\_N.  
DR Pfam: PF00488; Muts\_C; 1.  
DR Pfam: PF01624; Muts\_N; 1.  
DR ProDom: PD001263; Muts\_C; 1.  
DR SMART: SM00534; Mutsac; 1.  
DR SMART: SM00533; Mutsd; 1.  
DR PROSITE: PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
KW DNA repair; ATP-binding; DNA-binding; Complete proteome.  
FT NP\_BIND 602 609 ATP (POTENTIAL).  
SQ SEQUENCE 851 AA; 95470 MW; 737D51C8DCEFAF0A CRC64;

Query Match 96.6%; Score 28; DB 1; Length 851;  
Best Local Similarity 80.0%; Pred. No. 73;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||:|  
Db 758 SYGII 762

RESULT 9  
MTS\_BACSU STANDARD; PRT; 858 AA.  
ID MTS\_BACSU  
AC P49849;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE DNA mismatch repair protein mutS.  
GN MTS.  
OS *Bacillus subtilis*.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC *Bacillus/staphylococcus* group; *Bacillus*.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Gineti F., Perego M., Albertini A.M., Galizzi A.;  
RT "Bacillus subtilis mutS mult operon: Identification, nucleotide  
RT sequence and mutagenesis.";  
RL Microbiology 142:2021-2029(1996).  
CC -1- FUNCTION: This protein is involved in the repair of mismatches in  
CC DNA. It is possible that it carries out the mismatch recognition  
CC step. This protein has a weak ATPase activity (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MTS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U27343; AAB19235.1; ALT\_INIT.  
DR EMBL: 299112; CAB13577.1; -  
DR Subtilist; BG11403; mutS.  
DR InterPro: IPR000432; Muts\_C.  
DR InterPro: IPR002863; Muts\_N.  
DR Pfam: PF00488; Muts\_C; 1.  
DR Pfam: PF01624; Muts\_N; 1.  
DR ProDom: PD001263; Muts\_C; 1.  
DR SMART: SM00534; Mutsac; 1.  
DR SMART: SM00533; Mutsd; 1.  
DR PROSITE: PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
KW DNA repair; ATP-binding; DNA-binding; Complete proteome.  
FT NP\_BIND 602 609 ATP (POTENTIAL).  
SQ SEQUENCE 858 AA; 97590 MW; 641BDCDBCA4BC5D CRC64;

Query Match 96.6%; Score 28; DB 1; Length 858;  
Best Local Similarity 80.0%; Pred. No. 74;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||:|  
Db 758 SYGII 762

RESULT 10  
RS15\_MYCLE STANDARD; PRT; 89 AA.  
ID RS15\_MYCLE  
AC O32967;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 30S ribosomal protein S15.  
GN RPSO OR M10853 OR M1CB22.28C.  
OS *Mycobacterium leprae*.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; *Mycobacteriaceae*; *Mycobacterium*.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN.  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,  
RA Holtroyd S., Hornsby T., Jags K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 409:1007-1011(2001).  
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING  
CC PROTEINS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z98741; CAB11393.1; -  
DR EMBL: AL583920; CAC31234.1; -

DR HSSP: P05766; 1A32.  
 DR Leptoma; ML0853; -;  
 DR InterPro: IPR000589; Ribosomal\_S15.  
 DR Pfam: PF00312; Ribosomal\_S15; 1.  
 DR PROSITE: PS00362; RIBOSOMAL\_S15; 1.  
 KW Ribosomal protein; rRNA-binding; Complete proteome.  
 SQ SEQUENCE 89 AA: 10348 MW: A175AC55204571EE CRC64;

Query Match 89.7%; Score 26; DB 1; Length 89;  
 Best Local Similarity 80.0%; Pred. No. 23;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 111:1  
 Db 14 SYGLH 18

RESULT 11  
 RS15\_MYCTU STANDARD: PRT; 89 AA.  
 AC O33327;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S15.  
 GN RPSO OR RV2785C OR MT2855 OR MTW002.50C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eissen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE 16S RIBOSOMAL RNA BINDING  
 CC PROTEINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE S15P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AL008967; CA15580.1; -;  
 DR EMBL: AE007112; AAK47174.1; -;  
 DR HSSP: P05766; 1A32.  
 DR TIGR: MT2855; -;

DR TubercuList: RV2785C; -;  
 DR InterPro: IPR000589; Ribosomal\_S15.  
 DR Pfam: PF00312; Ribosomal\_S15; 1.  
 DR PROSITE: PS00362; RIBOSOMAL\_S15; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 89 AA: 10475 MW: 6DD2577898F97D7 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 89;  
 Best Local Similarity 80.0%; Pred. No. 23;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 111:1  
 Db 14 SYGLH 18

RESULT 12  
 Y965\_MYCTU STANDARD: PRT; 139 AA.  
 ID Y965\_MYCTU  
 AC P71545;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 14.5 kDa protein RV0965C.  
 GN RV0965C OR MT0993 OR MTCY10D7.09.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eissen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO M.TUBERCULOSIS RV2798C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z79700; CA01986.1; -;  
 DR EMBL: AE006984; AAK45242.1; ALT\_INIT.  
 DR TIGR: MT0993; -;  
 DR TubercuList: RV0965C; -;  
 DR Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 139 AA: 14479 MW: D6DAF1B4D1A1332 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 139;  
Best Local Similarity 80.0%; Pred. No. 35;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
DB 99 SYGVH 103

RESULT 13

CU08\_LOCM1 STANDARD; PRT; 148 AA.  
ID CU08\_LOCM1  
AC P11734;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Cuticle protein 8 (LM-8) (LM-ACP 8).  
OS Locusta migratoria (Migratory locust).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Orthoptera; Orthoptera; Caelifera;  
OC Acridomorpha; Acridoidea; Acrididae; Locusta.  
OX NCBI\_TaxID=7004;  
RN [1]  
RP SEQUENCE.

RA MEDLINE=90073593; PubMed=2590176;  
RA Klarskov K., Hoejrup P., Andersen S.O., Roepstorff P.;  
RT "Plasma-desorption mass spectrometry as an aid in protein sequence  
determination. Application of the method on a cuticular protein from  
the migratory locust (Locusta migratoria).";  
RL Biochem. J. 262:923-930(1989).  
RN [2]

RP SEQUENCE OF 1-56.  
RX MEDLINE=86108304; PubMed=3943519;  
RA Hoejrup P., Andersen S.O., Roepstorff P.;  
RT "Isolation, characterization, and N-terminal sequence studies of  
cuticular proteins from the migratory locust, Locusta migratoria.";  
RL Eur. J. Biochem. 154:153-159(1986).  
CC -1- FUNCTION: COMPONENT OF THE CUTICLE OF MIGRATORY LOCUST WHICH  
CONTAINS MORE THAN 100 DIFFERENT STRUCTURAL PROTEINS.  
CC -1- DOMAIN: THE TETRAPEPTIDE (A-A-P-[AV]) REPEATS FOUND THROUGHOUT THE  
PROTEIN ARE ALSO PRESENT IN MANY PROTEINS CONSTITUTING THE  
PROTECTIVE ENVELOPE OF OTHER SPECIES.  
CC -1- SIMILARITY: CONTAINS 1 CUTICLE CONSENSUS DOMAIN.

DR PIR; B24802; B24802.  
DR PIR; S05638; S05638.  
DR InterPro; IPR000618; Insect\_cuticle.  
DR Pfam; PF00379; Insect\_cuticle; 1.  
DR PRINTS; PR00947; CUTICLE.  
DR PROSITE; PS00233; CUTICLE; 1.  
KM Structural protein; Cuticle; Repeat.

FT REPEAT 16 19 1.  
FT REPEAT 22 22 2.  
FT REPEAT 28 31 3.  
FT REPEAT 37 40 4.  
FT REPEAT 44 47 5.  
SQ SEQUENCE 148 AA; 15224 MW; C6EADC27C593ACE6 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 148;  
Best Local Similarity 80.0%; Pred. No. 37;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
DB 64 NYGVH 68

RESULT 14

3MGL\_RHIME STANDARD; PRT; 185 AA.  
ID 3MGL\_RHIME  
AC 092TTL;  
DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT Putative 3-methyladenine DNA glycosylase (EC 3.2.2.-).  
DE Purative 3-methyladenine DNA glycosylase (EC 3.2.2.-).  
GN B01416 OR S0B20709.

OS Rhizobium meliloti (Sinorhizobium meliloti).  
OG Plasmid pSymb (megaplasmid 2).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396508; PubMed=11481431;  
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
RA Golding B., Puehler A.;  
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-  
fixing endosymbiont Sinorhizobium meliloti.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
CC -1- SIMILARITY: BELONGS TO THE DNA GLYCOSYLASE MFG FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AL603647; CAC49816.1; -  
DR Hypothetical protein; DNA repair; Hydrolase; Plasmid;  
KW Complete proteome.  
SQ SEQUENCE 185 AA; 19974 MW; 7DE6283605FA4A11 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 185;  
Best Local Similarity 80.0%; Pred. No. 47;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
DB 74 SYGLH 78

RESULT 15

GRP3\_DAUCA STANDARD; PRT; 195 AA.  
ID GRP3\_DAUCA  
AC P37705;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Glycine rich protein A3.  
OS Daucus carota (Carrot).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.  
OX NCBI\_TaxID=4039;

RN [1]  
RP SEQUENCE FROM N.A.  
RP Schrader S., Kaldenhoff R., Richter G.;  
RA Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X72383; CAAS1076.1; -  
DR PIR; S32123; S32123.  
SQ SEQUENCE 195 AA; 19099 MW; 886078886107DAA5 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 195;  
Best Local Similarity 80.0%; Pred. No. 49;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 SYGVH 5  
:1111  
Db 115 AYGVH 119

Search completed: August 20, 2002, 13:17:51  
Job time: 260 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:13:26 ; Search time 73.7 Seconds  
(Without alignments)  
11.736 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_28\_32  
Perfect score: 29  
Sequence: 1 SYGVH 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_oranelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	86	9 Q9ZX39	Q9ZX39 mycobacteri
2	29	100.0	188	11 Q9DB11	Q9DB11 mus musculu
3	29	100.0	221	12 Q91416	Q91416 sulfolobus
4	29	100.0	240	16 Q99088	Q99088 staphylococ
5	29	100.0	285	2 Q9X5X8	Q9X5X8 pseudomonas
6	29	100.0	285	16 Q51514	Q51514 pseudomonas
7	29	100.0	288	16 Q99013	Q99013 staphylococ
8	29	100.0	367	2 Q87084	Q87084 helicobacte
9	29	100.0	352	5 Q9BL53	Q9BL53 leishmania
10	29	100.0	397	17 Q980C6	Q980C6 sulfolobus
11	29	100.0	461	2 Q9ANAS	Q9ANAS bradyrhizob
12	29	100.0	490	10 Q9FSS8	Q9FSS8 oryza sativ
13	29	100.0	631	5 Q97340	Q97340 photonis va
14	29	100.0	657	10 Q9STF3	Q9STF3 arabidopsis
15	29	100.0	703	10 Q9ST08	Q9ST08 brassica ca
16	29	100.0	806	16 Q9ZM29	Q9ZM29 helicobacte

17	29	100.0	821	5 Q9VL46	Q9VL46 drosophila
18	29	100.0	831	16 Q24906	Q24906 helicobacte
19	29	100.0	871	17 Q9HSM2	Q9HSM2 halobacteri
20	29	100.0	1124	5 Q9VYN6	Q9VYN6 drosophila
21	29	100.0	1274	3 Q9UT05	Q9UT05 schizosach
22	29	100.0	4881	2 Q9S0R3	Q9S0R3 streptomyce
23	28	96.6	82	7 Q9XR7	Q9XR7 rattus fusc
24	28	96.6	126	2 Q48988	Q48988 mycoplasma
25	28	96.6	133	12 Q41853	Q41853 duvenhage v
26	28	96.6	139	16 Q9KAF7	Q9KAF7 bacillus h
27	28	96.6	176	16 Q98PA3	Q98PA3 rhizobium l
28	28	96.6	200	16 Q98IJ3	Q98IJ3 rhizobium l
29	28	96.6	289	16 Q9PHY0	Q9PHY0 campylobact
30	28	96.6	324	16 Q53481	Q53481 mycobacteri
31	28	96.6	434	2 Q9Z6C0	Q9Z6C0 enterobacte
32	28	96.6	450	12 Q82994	Q82994 lagos bat v
33	28	96.6	451	12 Q66453	Q66453 duvenhage v
34	28	96.6	600	10 Q41759	Q41759 zea mays (m
35	28	96.6	672	10 Q9LPA0	Q9LPA0 arabidopsis
36	28	96.6	804	10 Q94HX0	Q94HX0 oryza sativ
37	28	96.6	818	3 Q43059	Q43059 schizosach
38	28	96.6	841	2 Q93R04	Q93R04 lactococcus
39	28	96.6	860	16 Q92BV3	Q92BV3 listeria in
40	28	96.6	861	5 Q9VKT8	Q9VKT8 drosophila
41	28	96.6	1268	10 Q9SDC5	Q9SDC5 oryza sativ
42	28	96.6	1408	10 Q9FMU0	Q9FMU0 oryza sativ
43	28	96.6	1470	10 Q9AUP7	Q9AUP7 oryza sativ
44	28	96.6	1572	5 Q44938	Q44938 haemonchus
45	26	89.7	33	11 Q9Q232	Q9Q232 mus musculu

#### ALIGNMENTS

RESULT 1  
Q9ZX39 Q9ZX39 PRELIMINARY; PRT; 86 AA.  
AC Q9ZX39;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
DE GP39.  
OS Mycobacteriophage TM4.  
OC Viruses.  
OX NCBI\_TaxID=68870;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20110038; PubMed=10645443;  
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;  
RT "Mycobacteriophage TM4: Genome structure and gene expression.";  
RL Tuber. Lung Dis. 79:63-73(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF068845; AAD17606.1; -  
SQ SEQUENCE 86 AA; 9303 MW; 77382ADCC082323 CRC64;

Query Match 100.0%; Score 29; DB 9; Length 86;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYGVH 5  
Db 76 SYGVH 80

RESULT 2  
Q9DB11 Q9DB11 PRELIMINARY; PRT; 188 AA.  
AC Q9DB11;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE 2010001M09RIK PROTEIN.  
GN 2010001M09RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Botfeill D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakanato N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilmink L.,  
RA Wyszew-Borls A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL; AK008016; BAB25410.1; -  
DR MGD; MGI:1917066; 2010001M09RIK.  
DR InterPro; IPR000866; ER\_target.  
DR PROSITE; PS00014; ER\_TARGET; UNKNOWN\_1.  
SQ SEQUENCE 188 AA; 20559 MW; C66F7BC3996A8278 CRC64;

Query Match 100.0%; Score 29; DB 11; Length 188;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYGVH 5  
Db 100 SYGVH 104

RESULT 3  
091416 PRELIMINARY; PRT; 221 AA.  
AC 091416;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 25.5 KDA PROTEIN.  
GN SIFV0046.  
OS Sulfolobus islandicus filamentous virus.  
OC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;  
OC Lipothirixvirus.  
OX NCBI\_TaxID=176106;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20130740; PubMed=10662621;  
RA Arnold H.P., Zillig W., Ziese U., Holz I., Crosby M., Uterback T.,  
RA Weidmann J.F., Umayam L.A., Telfera K., Kristjanson J.K., Klenk H.P.,  
RA Nelson K.E., Fraser C.M.;  
RT "A novel lipothirixvirus, SIFV, of the extremely thermophilic  
KT crenarchaeon Sulfolobus.";  
RL Virology 267:252-266(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Arnold H.P., Zillig W., Ziese U., Holz I., Crosby M., Uterback T.,

RA Weidmann J.F., Umayam L.A., Telfera K., Kristjanson J.K., Klenk H.P.,  
RA Nelson K.E., Fraser C.M.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF440571; AAL27755.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 221 AA; 25486 MW; C278CC99CDF11D15 CRC64;

Query Match 100.0%; Score 29; DB 12; Length 221;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYGVH 5  
Db 69 SYGVH 73

RESULT 4  
099088 PRELIMINARY; PRT; 240 AA.  
ID 099088  
AC 099088;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE D1HYR0DIPICOLINATE REDUCTASE.  
GN DABP OR SA1228 OR SAV1396.  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain Mu50).  
OC Bacteria; Firmicutes; Bacilli; Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158879, 158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshiyama A.,  
RA Mizutani Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiratsuka K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
KT aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL; AP003133; BAB42488.1; -  
DR EMBL; AP003362; BAB57558.1; -  
DR HSSP; P04036; 1DRW.  
DR InterPro; IPR000846; DABP.  
DR Pfam; PF01113; DABP; 1.  
DR ProDom; PD004105; DABP; 1.  
KW Complete proteome.  
SQ SEQUENCE 240 AA; 26726 MW; F910FBD3DDBF4B8D CRC64;

Query Match 100.0%; Score 29; DB 16; Length 240;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SYGVH 5  
Db 107 SYGVH 111

RESULT 5  
09X5X8 PRELIMINARY; PRT; 285 AA.  
ID 09X5X8  
AC 09X5X8;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE PHA DEPOLYMERASE.  
GN PHA2.

OS Pseudomonas resinovorans.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=53412;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL B-2649;  
RA SOLAIMAN D.K.Y.;  
RT "PCR cloning of Pseudomonas resinovorans polyhydroxyalkanoate  
biosynthesis genes and expression in Escherichia coli.";  
RL Biotechnol. Lett. 22:789-794(2000).  
DR EMBL: AF129396; AAD26366.1; -  
DR InterPro: IPR000073; Abhydrolase.  
DR InterPro: IPR003088; AB\_hydrolase.  
DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
DR Pfam: PF00561; abhydrolase; 1.  
DR PRINTS: PR00111; ABHYDROLASE.  
DR PROSITE: PS00120; LIPASE\_SER; UNKNOWN\_1.  
SQ SEQUENCE 285 AA; 31463 MW; F23856753C801B0E CRC64;

Query Match 100.0%; Score 29; DB 2; Length 285;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 154 SYGVH 158

RESULT 6  
ID 051514 PRELIMINARY; PRT; 285 AA.  
AC 051514;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE PPA-DEPOLYMERASE (POLY(3-HYDROXYALKANOIC ACID) DEPOLYMERASE).  
GN PHAD OR PA5057.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RA Steinhuechel A.;  
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=93011120; Pubmed-1396693;  
RA Timm A., Steinhuechel A.;  
RT "Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid)  
gene locus of Pseudomonas aeruginosa PA01.";  
RL Eur. J. Biochem. 209:15-30(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; Pubmed-10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laribig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.  
DR EMBL: X66592; CAA47152.1; -  
DR EMBL: AE004919; AAG08442.1; -.

DR InterPro: IPR000073; Abhydrolase.  
DR InterPro: IPR003088; AB\_hydrolase.  
DR InterPro: IPR000379; Est\_lip\_thioest\_actsite.  
DR Pfam: PF00561; abhydrolase; 1.  
DR PRINTS: PR00111; ABHYDROLASE.  
KW Complete proteome.  
SQ SEQUENCE 285 AA; 31470 MW; C384C7D6435D5C6C CRC64;

Query Match 100.0%; Score 29; DB 16; Length 285;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 60 SYGVH 64

RESULT 7  
ID 099013 PRELIMINARY; PRT; 288 AA.  
AC 099013;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE SA1132 PROTEIN (HYPOTHETICAL PROTEIN SAV1290).  
GN SA1132 OR SAV1290.  
OS Staphylococcus aureus (strain N315), and  
OC Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=158879, 158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);  
RX MEDLINE=21311952; Pubmed-11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,  
RA Kanemori M., Matsunara H., Maruyama A., Murakami H., Hoshoyama A.,  
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kubara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus.";  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003133; BAB42385.1; -  
DR EMBL: AP003361; BAB57452.1; -  
DR InterPro: IPR000399; TPP\_enzyme.  
DR Pfam: PF02775; TPP\_enzymes\_C; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 288 AA; 31331 MW; 3A2AD0882C7FBCD4 CRC64;

Query Match 100.0%; Score 29; DB 16; Length 288;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
|||||  
Db 60 SYGVH 64

RESULT 8  
ID 087084 PRELIMINARY; PRT; 367 AA.  
AC 087084;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE CELL DIVISION-RELATED GENE.  
GN CDRA.  
OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HPK5;  
RX MEDLINE=98422487; PubMed=9748467;  
RA Taneuchi H., Shirai M., Akada J.K., Tsuda M., Nakazawa T.;  
RT "Nucleotide sequence and characterization of cdra, a cell division-  
related gene of Helicobacter pylori.";  
RL J. Bacteriol. 180:5263-5268(1998).  
DR EMBL: AB003309; BAA33499.1; -;  
DR InterPro: IPR002543; FtsK\_SpoIIIE.  
DR Pfam: PF01580; FtsK\_SpoIIIE; 1.  
SQ SEQUENCE 367 AA; 42500 MW; C6D43C01D6DB1434 CRC64;  
  
Query Match 100.0%; Score 29; DB 2; Length 367;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYGVH 5  
DB 321 SYGVH 325  
  
RESULT 9  
Q9BLS3 PRELIMINARY; PRT; 392 AA.  
ID Q9BLS3  
AC Q9BLS3;  
DT 01-JUN-2001 (TRENBLREL. 17, Created)  
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)  
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)  
DE HYPOTHEITICAL 41.1 KDA PROTEIN.  
GN L6071.08.  
OS Leishmania major.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RA Zimmermann W., Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;  
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FRIEDLIN;  
RX Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
RA Smith D.F.;  
RT "A physical map of the Leishmania major Friedlin genome.";  
RT Genome Res. 8:135-145(1998).  
DR EMBL: AL583933; CAC32267.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 392 AA; 41087 MW; 4A7DC7FC291785E3 CRC64;  
  
Query Match 100.0%; Score 29; DB 5; Length 392;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYGVH 5  
DB 231 SYGVH 235  
  
RESULT 10  
Q980C6 PRELIMINARY; PRT; 397 AA.  
ID Q980C6  
AC Q980C6;  
DT 01-OCT-2001 (TRENBLREL. 18, Created)  
DT 01-OCT-2001 (TRENBLREL. 18, Last sequence update)  
DT 01-OCT-2001 (TRENBLREL. 18, Last annotation update)  
DE HYPOTHEITICAL PROTEIN SSC00390.

GN SSC00390.  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
OX NCBI\_TaxID=2287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=11427726;  
RA She Q., Singh R.K., Contaltonieri F., Zivanovic Y., Allard G.,  
RA Aweez M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,  
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
DR EMBL: AE006672; AKA0717.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 397 AA; 41545 MW; 01C2618BAD7C2DA CRC64;  
  
Query Match 100.0%; Score 29; DB 17; Length 397;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYGVH 5  
DB 237 SYGVH 241  
  
RESULT 11  
Q9ANAS PRELIMINARY; PRT; 461 AA.  
ID Q9ANAS  
AC Q9ANAS;  
DT 01-JUN-2001 (TRENBLREL. 17, Created)  
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)  
DT 01-OCT-2001 (TRENBLREL. 18, Last annotation update)  
DE ID397.  
GN ID397.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Bradyrhizobium group; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=110SPC4;  
RX MEDLINE=21101824; PubMed=1157954;  
RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,  
RA Hennecke H.;  
RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb  
RT DNA region of the Bradyrhizobium japonicum chromosome.";  
RT J. Bacteriol. 183:1405-1412(2001).  
DR EMBL: AF322012; AAG60868.1; -;  
DR InterPro: IPR000567; SBP\_bac\_1.  
DR Pfam: PF01547; SBP\_bacterial\_1; 1.  
SQ SEQUENCE 461 AA; 50616 MW; 5043D981CC62A771 CRC64;  
  
Query Match 100.0%; Score 29; DB 2; Length 461;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SYGVH 5  
DB 240 SYGVH 244  
  
RESULT 12  
Q9FSS8 PRELIMINARY; PRT; 490 AA.  
ID Q9FSS8  
AC Q9FSS8;  
DT 01-MAR-2001 (TRENBLREL. 16, Created)



DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE PTNATIVE WALL-ASSOCIATED KINASE 2.  
 GN H0212B02.2.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Han B., Zhou B., Feng Q., Huang Y.C., Chen Z.H., Li Y., Zhu J.J.,  
 RA Tang Y.S., Zhao Q., Liu Y.L., Mu J., Yu Z., Fan D.L., Chen L.,  
 RA Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Zhu J., Liu X.H., Hu X.,  
 RA Lei H.Y., Zhang Y.J., Wang R., Li C., Lu Y., Chen X.C., Zhang Y.,  
 RA Hu H., Jia P.X., Li T., Qian Y.M., Ying K., Hong G.F.,  
 RA "Oryza sativa indica (Guangluai4) genomic DNA, chromosome4, BAC  
 RT clone:H0212B02."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL442007; CAC09346.1; -  
 KW Kinase.  
 SO SEQUENCE 490 AA; 52039 MW; 2F1301B08BA26EFE CRC64;

Query Match 100.0%; Score 29; DB 10; Length 490;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SYGVH 5  
 |||||  
 DB 186 SYGVH 190

RESULT 13  
 097340  
 ID 097340 PRELIMINARY; PRT; 631 AA.  
 AC 097340;  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE CYTOPLASMIC INTERMEDIATE FILAMENT PROTEIN.  
 OS Photons vanconverensis.  
 OC Eukaryota; Metazoa; Brachyopoda; Photoniiformes; Photons.  
 OX NCBI\_TaxID=34521;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-99065768; Pubmed-9847417;  
 RA Eder A., Riemer D., Boveneschulte M., Weber K.;  
 RT "Molecular Phylogeny of metazoan intermediate filament proteins."  
 RL J. Mol. Evol. 47:751-762(1998).  
 DR EMBL; AJ004936; CAB38181.1; -  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR001322; IF\_tail.  
 DR Pfam: PF00038; Filament; 1.  
 DR Pfam: PF00932; IF\_tail; 1.  
 DR PROSITE: PS00226; IF; UNKNOWN\_1.  
 SO SEQUENCE 631 AA; 71028 MW; 64856BEF7D82379 CRC64;

Query Match 100.0%; Score 29; DB 5; Length 631;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 |||||  
 DB 71 SYGVH 75

RESULT 14  
 09STF3  
 ID 09STF3 PRELIMINARY; PRT; 657 AA.  
 AC 09STF3;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 74.3 KDA PROTEIN.  
 GN T6H20.180.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Choisme N., Robert C., Brothier P., Wincker P., Catolico L.,  
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,  
 RA Lemcke K., Schueller C., Queller F., Salanoubat M.;  
 RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL096859; CAB51186.1; -  
 DR InterPro: IPR002885; PPR.  
 DR Pfam: PF01535; PPR; 6.  
 DR Hypothetical protein.  
 KW  
 SO SEQUENCE 657 AA; 74340 MW; 8257019E45CCDFC8 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 657;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 SYGVH 5  
 |||||  
 DB 360 SYGVH 364

RESULT 15  
 09ST08  
 ID 09ST08 PRELIMINARY; PRT; 703 AA.  
 AC 09ST08;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE TRANSPOSON-LIKE ORF PROTEIN.  
 GN TRANSPOSON-LIKE ORF.  
 OS Brassica campestris (Field mustard).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3711;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-99403015; Pubmed-10471721;  
 RA Suzuki G., Kai N., Hirose T., Fukui K., Nishio T., Takayama S.,  
 RA Isogai A., Watanabe M., Hinata K.;  
 RT "Genomic organization of the S locus: identification and  
 RT characterization of genes in SlG/SRK region of S9 haplotype of  
 RT Brassica campestris (syn. rapa).";  
 RL Genetics 153:391-400(1999).  
 DR EMBL; AB022082; BAB85462.1; -  
 DR InterPro: IPR004242; Transposase\_21.  
 DR Pfam: PF02992; Transposase\_21; 1.  
 SO SEQUENCE 703 AA; 81881 MW; AC46541BEEC83C79 CRC64;

Query Match 100.0%; Score 29; DB 10; Length 703;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYGVH 5  
 |||||  
 DB 546 SYGVH 550

Aug 20 13:05:57 2002

us-09-824-286-3\_copy\_28\_32.rspt

Page 6

Search completed: August 20, 2002, 13:17:23  
Job time: 237 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - .protein search, using sw model

Run on: August 20, 2002, 13:15:09 ; Search time 94.22 Seconds  
(without alignments)  
17.683 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_47\_61  
Perfect score: 80  
Sequence: 1 VIMAGSGTMYNSALM 15

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: \_A\_Geneseq\_032802.\*  
2: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
3: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
4: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
5: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
6: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
7: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
8: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
9: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
10: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
11: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
12: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
13: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
14: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
15: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
16: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
17: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
18: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
19: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
20: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
21: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
22: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
23: /SIDSL1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	16	AAAB1965	Ganglioside GD2 sp
2	80	100.0	16	AAAG3989	Monoclonal antibody
3	80	100.0	112	AAAW1648	Monoclonal antibody
4	80	100.0	113	AAAR21268	Murine VH group 2
5	80	100.0	116	AAAR07322	Murine domain of anti
6	80	100.0	116	AAAW14450	Monoclonal antibody
7	80	100.0	116	AAAR99877	Monoclonal antibody
8	80	100.0	118	AAAB1970	Ganglioside GD2 sp
9	80	100.0	118	AAAB1973	Ganglioside GD2 sp
10	80	100.0	120	AAAG3986	Amino acid sequenc
11	80	100.0	137	AAAB1975	Ganglioside GD2 sp

12	80	100.0	222	14	AAAR32843	VH NQ2/12.4-Vk NQ1
13	80	100.0	235	14	AAAR32840	VH NQ2/12.4-Vk NQ1
14	80	100.0	269	13	AAAR32569	Fusion protein enc
15	80	100.0	581	22	AAAB1972	Ganglioside GD2 sp
16	77	96.2	142	22	AAAG65520	Mouse antibody 26
17	77	96.2	142	22	AAAG65523	Humanised anti-CTL
18	77	96.2	239	14	AAAR43679	Single chain polyp
19	77	96.2	239	17	AAAW02191	18-2-3/TRR202' sin
20	77	96.2	239	17	AAAR96649	Single chain bindi
21	77	96.2	241	11	AAAR06482	18-2-3-/TRR202' .
22	77	96.2	242	11	AAAR06483	18-2-3-/TRR159. A
23	77	96.2	242	14	AAAR43680	Single chain polyp
24	77	96.2	242	17	AAAW02192	18-2-3/TRR159 singl
25	77	96.2	242	17	AAAR96650	Single chain bindi
26	76	95.0	272	18	AAAW16688	Murine anti-human
27	74	92.5	140	18	AAAW2538	Murine anti-human
28	71	88.8	29	18	AAAW13989	Single chain antib
29	71	88.8	29	18	AAAW13988	Single chain antib
30	71	88.8	29	18	AAAW13987	Single chain antib
31	71	88.8	29	18	AAAW13986	Single chain antib
32	71	88.8	476	22	AAAB49243	Chimeric 4H6 anti-
33	70	87.5	119	21	AAAY0818	260F9 hybridoma VL
34	69.5	86.9	115	18	AAAW04595	Anti-DNA antibody
35	69.5	86.9	123	18	AAAW07438	Anti-DNA antibody
36	66	82.5	231	18	AAAW27090	Mouse monoclonal a
37	65	81.2	16	17	AAAR8481	Anti-IL-5 Mab heav
38	65	81.2	111	17	AAAR8494	Mab 2E3 heavy cha
39	65	81.2	111	17	AAAR8496	Mab 2E3 heavy cha
40	65	81.2	111	19	AAAW42453	Mouse anti-human I
41	65	81.2	111	19	AAAW42455	Mouse anti-human I
42	65	81.2	119	17	AAAR8478	Mab 2B6 heavy cha
43	65	81.2	119	17	AAAR8488	Humanised 2B6 anti
44	65	81.2	119	17	AAAR8492	HEM humanised 2B6
45	65	81.2	119	19	AAAW42465	Interleukin-5 huma

#### ALIGNMENTS

RESULT 1	AAAB1965	standard; Peptide: 16 AA.
ID	AAAB1965	
AC	AAAB1965;	
XX	03-JUL-2001	(first entry)
DE	Ganglioside GD2 specific antibody related peptide SEQ ID NO: 4.	
XX	Ganglioside GD2; complementation determining region; CDR; antibody;	
KW	mouse; cancer.	
XX		
OS	Mus musculus.	
XX		
PN	WO200123573-A1.	
XX		
PD	05-APR-2001.	
XX		
PF	29-SEP-2000; 2000WO-JP06773.	
XX		
PR	30-SEP-1999; 99JP-0278290.	
XX		
PA	(KYOW ) KYOWA HAKKO KOGYO KK.	
XX		
PI	Hanai N, Shtata K, Nakamura K, Niwa R;	
XX		
DR	WPI; 2001-266163/27.	
XX		
PT	Human type complementation-determining domain transplanted antibody and	
PT	derivatives against ganglioside GD2, useful in diagnosis and therapy of	
PT	e.g. tumors, has low antigenicity, little side effects but potent	
PT	activity in cancer	
XX		

PS Claim 5; Page 99; 123pp: Japanese.

CC The present invention describes an antibody, which can react specifically

CC with ganglioside GD2, and is transplanted with a human type

CC complementation-determining domain (CDR), or its fragments. The antibody

CC and its derivatives are useful in diagnosis and therapy of tumours,

CC particularly cancer diagnosis. The present sequence is a peptide

CC used in the exemplification of the invention.

XX Sequence 16 AA:

SO

Cy 1 VIMAGSTNYNSALM 15  
 |||||  
 1 vimagstnynsalm 15

Db

RESULT 2  
 AAG63989  
 ID AAG63989 standard; peptide: 16 AA.  
 AC AAG63989;  
 XX  
 XX 26-NOV-2001 (first entry)  
 DE Complementarity determining region of heavy chain of antibody 2C4.  
 KW Monoclonal antibody 2C4; sialoadhesin factor-2; SAF-2; allergic rhinitis;  
 KW allergy; asthma; anemia; eczema; lymphoma; systemic mastocytosis;  
 KW leukemia; eosinophil.  
 XX  
 XX Mus sp.  
 OS  
 XX WO200166126-A1.  
 XX  
 XX 13-SEP-2001.  
 PD  
 XX 05-MAR-2001; 2001WO-US07193.  
 PF  
 XX 07-MAR-2000; 2000US-0187595.  
 BR  
 XX (SMIR ) SMITHKLINE BEECHAM CORP.  
 PA (SMIR ) SMITHKLINE BEECHAM PLC.  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 XX  
 PI Abrahamson JA, Bochner B, Erickson-Miller CL, Kikly KK;  
 PI Schleimer R;  
 DR WPI; 2001-570749/64.  
 XX  
 XX Novel monoclonal antibody specific for human sialoadhesin factor-2 for  
 PT diagnosis, prevention, treatment of allergy, asthma, eczema or diseases  
 PT such as lymphoma, leukemia or systemic mastocytosis, in a mammal -  
 XX  
 PS Claim 10; Page 33; 35pp: English.

CC AAG63989-90 represent the complementarity determining regions (CDRs)

CC of the heavy chain variable region of murine monoclonal antibody 2C4.

CC This antibody binds to human sialoadhesin factor-2 (SAF-2). The

CC antibody is useful for treating or preventing allergic rhinitis,

CC allergies, asthma, anemia, eczema or diseases such as lymphoma,

CC leukemia or systemic mastocytosis in a mammal. It is also useful for

CC detecting the presence of a cell, especially eosinophil in a sample,

CC by detecting binding of the antibody to SAF-2. The antibody can be

CC coupled to toxins, antiproliferative drugs or radionuclides to

CC kill cells in areas of excessive SAF-2 expression.

XX Sequence 16 AA:

SO

Query Match 100.0%; Score 80; DB 22; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 VIMAGSTNYNSALM 15  
 |||||  
 1 vimagstnynsalm 15

Db

RESULT 3  
 AAW31648  
 ID AAW31648 standard; Protein: 112 AA.  
 AC AAW31648;  
 XX  
 XX 21-MAY-1998 (first entry)  
 DE Monoclonal antibody CP.B8 heavy chain variable region.  
 KW Cytokine receptor; gamma common chain; gc chain; human;  
 KW blocking agent; monoclonal antibody; CP.B8; immunological disease;  
 KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;  
 KW insulin-dependent diabetes; inflammatory bowel disease;  
 KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;  
 KW graft versus host disease; psoriasis; immunosuppressive; therapy;  
 KW complementarity determining region; CDR.  
 XX  
 XX Mus musculus.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT Region 28..32  
 FT /note= "CDR1"  
 FT Region 47..61  
 FT /note= "CDR2"  
 FT Region 95..104  
 FT /note= "CDR3"  
 XX  
 XX WO9743416-A1.  
 XX  
 XX 20-NOV-1997.  
 PD  
 XX 09-MAY-1997; 97WO-US07870.  
 PF  
 XX 10-MAY-1996; 96US-0017466.  
 PR  
 XX (BIOJ ) BIOGEN INC.  
 PA  
 XX Benjamin CD, Burkly LC, Hession C, Whitty A;  
 PI  
 XX WPI; 1998-008885/01.  
 DR N-PSDB; AAT97441.  
 DR  
 XX  
 XX Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell energy for  
 PT treatment of immunological diseases  
 XX  
 PS Claim 23; Page 81-82; 111pp: English.

CC This polypeptide comprises the heavy chain variable region (VH) of

CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma

CC deposited as ATCC 12107, and which is specific for the gamma

CC constant (gc) chain (see AAW31646) of human cytokine receptors. The

CC invention provides compositions and methods for inhibiting cytokine

CC signalling using gc chain blocking agents for the treatment of

CC immunological diseases such as myasthenia gravis, rheumatoid

CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,

CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,

CC allergy, asthma, parasitic infection, graft vs. host disease or

CC psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab

CC fragment and an antibody having a light chain variable region

CC CDR selected from those of CP.B8 VH or a heavy chain variable

CC region CDR selected from those of CP.B8 VL (see AAW31647).

```

XX SQ Sequence 112 AA;
Query Match 100.0%; Score 80; DB 19; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VIMAGSTNYNSALM 15
   |||
DB 47 vlmagstnynsalm 61

RESULT 4
AAR21268
ID AAR21268 standard; Protein; 113 AA.
XX AC AAR21268;
XX DT 21-MAY-1992 (first entry)
XX DE Murine VH group 2 chain E specific for phox.
XX KW Fg; bacteriophage; gene III; filamentous; phagemid; capsid; coat;
XX KW pilus; g3p; binding; adsorption; gene VIII; diverse repertoire;
XX KW specific binding pairs; replicable genetic display package.
XX OS Synthetic.
XX FH Key
XX FT Binding-site Location/Qualifiers
XX FT /label= CDR1 31..35
XX FT Binding-site 50..65
XX FT /label= CDR2 98..102
XX FT Binding-site /label= CDR3
XX FT /note= "D/N-X-G-X-X motif"
XX FT
XX PN WO9201047-A.
XX PD 23-JAN-1992.
XX PF 10-JUL-1991; 91WO-GB01134.
XX PR 15-MAY-1991; 91GB-0010549.
XX PR 10-JUL-1990; 90GB-0015198.
XX PR 19-OCT-1990; 90GB-0022845.
XX PR 12-NOV-1990; 90GB-0024503.
XX PR 06-MAR-1991; 91GB-0004744.
XX PA (CAMP-) CAMBRIDGE ANTIBODY.
XX PA (MEDI-) MED RES COUNCIL.
XX PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;
XX PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;
XX PI Winter GP, Bonfert TP;
XX DR WPI; 1992-056862/07.
XX PT Producing members of specific binding pairs - by expression in
XX PT recombinant host cells with a secreting replicable genetic
XX PT display package.
XX PS Example 21; Fig 24; 209pp; English.
XX CC The VH sequence is one of eight (AAR21264-71) found to be expressed
XX CC from a single chain Fv library from an immunised mouse. The
XX CC library produces a diverse repertoire of antibody fragments specific
XX CC for 2-phenyl-5-oxazolone (phox). It was prepd. using cDNA generated
XX CC from mRNA from mice immunised with phox coupled to chicken serum
XX CC albumin. The VH and VL kappa sequences were separately amplified by
XX CC PCR (see AAQ23474-84) and ligated into fdCAT2 (see AAQ23463) for
XX CC expression on the phage surface as fusions with gene III. The

```

```

CC resulting library of clones was diverse. Twenty three hapten binding
CC clones were sequenced revealing the eight different VH genes (A-H)
CC in a variety of pairings with seven different Vk genes (a-g) (see
CC AAR21264-92). Nearly all the VH genes belonged to gp 1, with only
CC one, "E", being of gp 2 (VHox1). Of the twenty three clones
CC sequenced, only one was of type "E". Most of the clones were Vk-d
CC combinations. The Kd of VH-B/Vk-d for phox-GABA was 10 nM. Only two
CC other combinations (of eleven tested) were found to have higher
CC values. This suggests that phage bearing scFv fragments having
CC weak affinities can be selected with antigen, probably due to the
CC avidity of the multiple antibody heads on the phage.
CC See also AAR21260-307, 309-311; AAR22450, 565-581.
XX SQ Sequence 113 AA;
Query Match 100.0%; Score 80; DB 13; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VIMAGSTNYNSALM 15
   .|||
DB 50 vlmagstnynsalm 64

RESULT 5
AAR07322
ID AAR07322 standard; protein; 116 AA.
XX AC AAR07322;
XX DT 22-JAN-1991 (first entry)
XX DE VH domain of antibody D against tumour-associated antigens.
XX KW Tumour-associated antigen; murine monoclonal antibody D;
XX KW Vibrio Cholerae neuraminidase-sensitive epitope; glioma;
XX KW meningioma; neurilemmoma; neuroblastoma; ganglioblastoma;
XX KW ganglioneuroma; diagnosis.
XX OS Mus musculus.
XX PN EP888914-A.
XX PD 26-SEP-1990.
XX PF 21-MAR-1990; 90EP-0105322.
XX PR 24-MAR-1989; 89DE-3909799.
XX PA (BEHM ) BEHRINGWERKE AG.
XX PI Bosslet K, Seemann G, Sedlacek HH;
XX PI WPI; 1990-291873/39.
XX DR N-PSDB; AAQ06229.
XX PT Monoclonal antibodies to tumour associated antigens - used for
XX PT diagnosis of malignant tumours etc.
XX PS Disclosure; Page 14; 18pp; German.
XX CC Antibody D is produced as described in EP-141079 and binds to
XX CC a vibrio cholerae neuraminidase-sensitive epitope on ganglioside Gp2
XX CC which occurs in gliomas, meningiomas, neurilemmomas,
XX CC neuroblastomas, ganglioblastomas and ganglioneuromas.
XX CC They are useful in tumour diagnosis and therapy.
XX CC See also AAQ06230 for VK of Mab D, AAQ07312-15 for Mab A and B and
XX CC AAQ06227-28 for Mab C.
XX SQ Sequence 116 AA;

```

Query Match 100.0%; Score 80; DB 11; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15  
 |||  
 DB 45 VIMAGSTNYNSALM 59

RESULT 6  
 AAM14490  
 ID AAM14490 standard; Protein: 116 AA.  
 AC AAM14490;  
 XX  
 XX 28-JAN-1997 (first entry)  
 DE Monoclonal antibody D VH.  
 heavy; light chain; monoclonal antibody; antigen 3; marker; melanoma;  
 permanent human tumour cell line; tumour-associated antigen; epitope;  
 gastrointestinal tumour; pancreatic carcinoma; diagnostic; therapeutic;  
 antigen 11; Vibrio cholera; neuraminidase-resistant; ganglioside GD2.  
 XX  
 OS Synthetic.  
 XX  
 XX EP727436-A1.  
 XX  
 XX 21-AUG-1996.  
 XX  
 XX 21-MAR-1990; 90EP-0105322.  
 XX  
 XX 24-MAR-1989; 89DE-3909799.  
 XX  
 XX (BEHW ) BEHRINGWERKE AG.  
 XX  
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 DR WPI: 1996-372836/38.  
 DR N-PSDB; AAT63507.  
 XX  
 PT Monoclonal antibody to tumour-associated antigen - useful as  
 PS gastrointestinal tumour marker  
 XX  
 PS Disclosure: Page 14; 19pp; German.  
 XX  
 CC AAM14490-91 are the heavy and light chains (respectively) of monoclonal  
 CC antibody (MAB) D. Mab D recognises Vibrio cholera  
 CC neuraminidase-resistant epitope of ganglioside GD2, from a human melanoma  
 CC cell line. Mabs A, B and C (see AAM14484-89) are mentioned in the  
 CC specification, but are not part of the claims. Mabs A and B recognise  
 CC antigens 3 and 11 resp., of a permanent human tumour cell line. Mab C  
 CC also recognises an epitope of a tumour-associated antigen. These antigens  
 CC occur at high concns. in the serum of patients with gastrointestinal  
 CC tumours, e.g. pancreatic carcinoma, and are thus useful as tumour markers  
 CC for diagnostic or therapeutic purposes.  
 XX  
 SQ Sequence 116 AA;

Query Match 100.0%; Score 80; DB 17; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15  
 |||  
 DB 45 VIMAGSTNYNSALM 59

RESULT 7  
 AAR99877  
 ID AAR99877 standard; Protein: 116 AA.  
 XX

AC AAR99877;  
 XX  
 DT 28-JAN-1997 (first entry)  
 XX  
 DE Monoclonal antibody D VH.  
 XX  
 XX Monoclonal antibody; MAB; epitope: tumour-associated antigen;  
 KW marker; antigen.  
 XX  
 OS Synthetic.  
 XX  
 XX EP727435-A1.  
 XX  
 XX 21-AUG-1996.  
 XX  
 XX 21-MAR-1990; 90EP-0105322.  
 XX  
 XX 24-MAR-1989; 89DE-3909799.  
 XX  
 XX (BEHW ) BEHRINGWERKE AG.  
 XX  
 PI Auerbach B, Bosslet K, Sedlacek H, Seemann G;  
 DR WPI: 1996-372835/38.  
 DR N-PSDB; AAT36665.  
 XX  
 PT Monoclonal antibody to tumour-associated antigen - useful as  
 PS gastrointestinal tumour marker  
 XX  
 PS Disclosure: Page 14; 19pp; German.  
 XX  
 CC MAB C (AAT36659-T36660) is a monoclonal antibody that recognises an  
 CC epitope of a tumour-associated antigen occurring at high concn. in  
 CC the serum of patients with gastrointestinal tumours, e.g. pancreatic  
 CC carcinoma, and is thus useful as a tumour marker for diagnostic or  
 CC therapeutic purposes.  
 CC  
 CC Mabs A, B and D are mentioned in the specification, but are not  
 CC part of the claims.  
 CC MAB A (AAT36661-T36662) recognises antigen 3 of permanent human  
 CC tumour cell line.  
 CC MAB B (AAT36663-T36664) recognises antigen 11 of permanent human  
 CC tumour cell line.  
 CC MAB D (AAT36665-T36666) recognises a Vibrio cholera neuraminidase-  
 CC resistant epitope of ganglioside GD2, from a human melanoma cell  
 CC line.  
 XX  
 SQ Sequence 116 AA;

Query Match 100.0%; Score 80; DB 17; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15  
 |||  
 DB 45 VIMAGSTNYNSALM 59

RESULT 8  
 AAB81970  
 ID AAB81970 standard; Protein: 118 AA.  
 AC AAB81970;  
 XX  
 XX 03-JUL-2001 (first entry)  
 DE Ganglioside GD2 specific antibody related protein SEQ ID NO: 13.  
 XX  
 XX Ganglioside; GD2; complementation determining region; CDR; antibody;  
 KW mouse; cancer.  
 XX  
 OS Synthetic.  
 XX

PN	WO200123573-A1.
XX	
PD	05-APR-2001.
XX	
XX	29-SEP-2000; 2000WO-JP06773.
PF	
XX	
PR	30-SEP-1999; 99JP-0278290.
XX	
PA	(KYOW ) KYOWA HAKKO KOGYO KK.
XX	
PI	Hanai N, Shitara K, Nakamura K, Niwa R;
XX	
DR	WPI; 2001-266163/27.
XX	
PT	Human type complementation-determining domain transplanted antibody and
PT	derivatives against ganglioside GD2, useful in diagnosis and therapy of
PT	e.g. tumours, has low antigenicity, little side effects but potent
PT	activity in cancer -
XX	
PS	Claim 10; Page 102-103; 123pp; Japanese.
XX	
XX	The present invention describes an antibody, which can react specifically
CC	with ganglioside GD2, and is transplanted with a human type
CC	complementation-determining domain (CDR), or its fragments. The antibody
CC	and its derivatives are useful in diagnosis and therapy of tumours,
CC	particularly cancer diagnosis. The present sequence is a protein
CC	used in the exemplification of the invention.
XX	
XX	Sequence 118 AA;
XQ	

Query Match	100.0%	Score 80	DB 22	Length 118
Best Local Similarity	100.0%	Pred. No. 12e-05		
Matches 15, Conservative 0		Mismatches 0	Indels 0	Gaps 0
OY	1	VIVAGGSTNYNSALM	15	
db	50	VIVAGGSTNYNSALM	64	

XX	RESULT	9
XX	AAB81973	ID AAB81973 standard; Protein; 118 AA.
XX	AC	AAB81973;
XX	DT	03-JUL-2001 (first entry)
XX	DE	Ganglioside GD2 specific antibody related protein SEQ ID NO: 32.
XX	KM	Ganglioside; GD2; complementation determining region; CDR; antibody;
KW	mouse; cancer.	
XX	OS	Mus musculus.
XX	PN	WO200123573-A1.
XX	PD	05-APR-2001.
XX	PF	29-SEP-2000; 2000WO-JP06773.
XX	PR	30-SEP-1999; 99JP-0278290.
XX	PA	(KYOW ) KYOMA HAKKO KOGYO KK.
XX	PI	Hanai N, Shitara K, Nakamura K, Niwa R;
XX	DR	WPI; 2001-266163/27.
PT	Human type complementation-determining domain transplanted antibody and derivatives against ganglioside GD2, useful in diagnosis and therapy of e.g. tumours, has low antigenicity, little side effects but potent	
PT	PT	

XX Claim 26; Page 115-116; 123pp; Japanese.  
PS  
XX  
XX The present invention describes an antibody, which can react specifically  
CC with ganglioside GD2, and is transplanted with a human type  
CC complementation-determining domain (CDR), or its fragments. The antibody  
CC and its derivatives are useful in diagnosis and therapy of tumors,  
CC particularly cancer diagnosis. The present sequence is a protein  
CC used in the exemplification of the invention.  
XX  
XX  
XX Sequence 118 AA;  
SO

Query Match	100.0%	Score 80	DB 22	Length 118
Best Local Similarity	100.0%	Pred. No. 1.2e-05		
Matches 15	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	VWAGGSTNYNSALM	15	
Db	50	VWAGGSTNYNSALM	64	

RESULT	10
AAG63986	
ID	AAG63986 standard; Protein; 120 AA
XX	
AC	AAG63986;

DT 26-NOV-2001 (first entry)

DE Amino acid sequence of heavy chain variable region of antibody 2C4.

KW Monoclonal antibody 2C4; sialoadhesin factor-2; SAE-2; allergic rhinitis;  
KW allergy; asthma; anemia; eczema; lymphoma; systemic mastocytosis;  
KW leukemia; eosinophil.

OS Mus sp.

PN WO200166126-A1.

PD 13-SEP-2001.

PF 05-MAR-2001; 2001WO-US07193

PR 07-MAR-2000; 2000US-0187595

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PA (UYJO ) UNIV JOHNS HOPKINS.

PI Abrahamson JA, Bochner B, Erickson-Miller CL, Kikly KK;

XX  
 11 10 9 8 7 6 5 4 3 2 1

DR N-PSDB; AAH78183.

Novel monoclonal antibody specific for human sialoadhesin factor-2 for

PT such as lymphoma, leukemia or systemic mastocytosis, in a mammal -

PS Claim 17; Fig 1; 35pp; English

The present sequence represents the heavy chain variable region of murine

factor-2 (SAF-2). The antibody is useful for treating or preventing

as lymphoma, leukemia or systemic mastocytosis in a mammal. It is also allergic rhinitis, allergies, asthma, anemia, eczema or diseases such

CC a sample by detecting the binding of the antibody to SAF-2. The antibody in  
CC use for detecting the presence of a cell, especially eosinophil in

CC kill cells in areas of excessive SAF-2 expression

SQ Sequence 120 AA;

Query Match 100.0%; Score 80; DB 22; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGGSTNYSALM 15  
 |||||||  
 DB 50 vimaggstnysalm 64

## RESULT 11

AAB81975  
 ID AAB81975 standard; Protein; 137 AA.

AC AAB81975;

DT 03-JUL-2001 (first entry)

DE Ganglioside GD2 specific antibody related protein #1.

KW Ganglioside; GD2; complementation determining region; CDR; antibody;  
 mouse; cancer.

OS Mus musculus.

PN WO200123573-A1.

PD 05-APR-2001.

PE 29-SEP-2000; 2000WO-JP06773.

PR 30-SEP-1999; 99JP-0278290.

PA (KYOW ) KYOWA HAKKO KOGYO KK.

PI Hanai N, Shitara K, Nakamura K, Niwa R;

DR WPI: 2001-266163/27.

DR N-PSDB; AAF86854.

PT Human type complementation-determining domain transplanted antibody and  
 PT derivatives against ganglioside GD2, useful in diagnosis and therapy of  
 PT e.g. tumours, has low antigenicity, little side effects but potent  
 PT activity in cancer -

XX Example 2; Page 96-97; 123pp; Japanese.

XX The present invention describes an antibody, which can react specifically  
 CC with ganglioside GD2, and is transplanted with a human type  
 CC complementation-determining domain (CDR), or its fragments. The antibody  
 CC and its derivatives are useful in diagnosis and therapy of tumours,  
 CC particularly cancer diagnosis. The present sequence is a protein  
 CC used in the exemplification of the invention.

SO Sequence 137 AA;

Query Match 100.0%; Score 80; DB 22; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGGSTNYSALM 15  
 |||||||  
 DB 69 vimaggstnysalm 83

## RESULT 12

AAR32843  
 ID AAR32843 standard; Protein; 222 AA.

AC AAR32843;

XX

DT 19-JUN-1993 (first entry)

XX VH NQ2/12.4-VK NQ10/12.5 linked peptide sequences #2.

DE Primer: human; immunoglobulin; Ig; variable region; VH; VL; Ck; JH;

KW lymphocyte; vector; soluble; antibody; phage; linker; back; VH3;

KW nested; In-cell PCR; cloning; polymorphic; TCR V; antiphenylloxazone;

OS hydridoma; NQ2/12.4; NQ10/12.5.

XX Synthetic.

XX Key

FT Region

FT Peptide

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

Query Match 100.0%; Score 80; DB 14; Length 222;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGGSTNYSALM 15  
 |||||||  
 DB 50 vimaggstnysalm 64

## RESULT 13

AAR32840  
 ID AAR32840 standard; Protein; 235 AA.

AC AAR32840;

XX



DT 19-JUN-1993 (first entry)  
XX  
DE VH NQ2/12.4-VK NQ10/12.5 linked peptide sequences.  
XX  
KW Primer; human; immunoglobulin; Ig; variable region; VH; VL; CK; JH;  
KW lymphocyte; vector; soluble; antibody; phage; linker; back; VH3;  
KW nested; in-cell PCR; cloning; polymorphic; TCR V; antiphenyloxazolone;  
KW hybridoma; NQ2/12.4; NQ10/12.5.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FH Region 1..113  
FT /label= VH\_NQ2/12.4  
FT 114..127  
FT /note= "Linker peptide"  
FT 128..235  
FT Region /label= V kappa\_NQ2/12.4  
FT  
XX  
PN WO9303151-A.  
XX  
PD 18-FEB-1993.  
XX  
PE 10-AUG-1992; 92WO-GB01483.  
XX  
PR 10-AUG-1991; 91GB-0017352.  
PR 11-JUN-1992; 92GB-0012419.  
XX  
PA (MED-) MEDICAL RES COUNCIL.  
XX  
PI Embleton MJ, Gorochov G, Jones PT, Winter GP;  
PI  
XX  
DR WPI; 1993-076508/09.  
DR N-PSDB; AAQ37459.  
XX  
PT Treatment of cell populations, partic. hybridomas - to link  
PT together copies of 2 or more non-contiguous DNA sequences to  
PT facilitate analysis  
PT  
XX  
PS Disclosure; Fig 2; 72pp; English.  
XX  
XX The sequences given in AAR32840-43 show the mature heavy chain VH  
CC domains and the VK light chain genes of the antiphenyloxazolone  
CC hybridomas NQ2/12.4 and NQ10/12.5 which have been linked via a linker  
CC peptide by in-cell PCR. The cDNA encoding these peptides was  
CC synthesised using forward primers annealing to the CK gene and the JH  
CC segment, followed by assembly with linker primers, VH back primers  
CC based on the VH3 leader sequence and a forward CK primer nested in  
CC respect to the primer used for cDNA. The assembled product within  
CC the cells is then amplified with nested primers annealing to the 5'  
CC end of the VH gene and the 3' end of the JK segment. In-cell PCR may  
CC be used to determine gene linkage analysis, particularly for the  
CC cloning of gene combinations that are polymorphic within a population  
CC of cells, such as the rearranged genes for Ig or TCR V regions.  
XX  
SQ Sequence 235 AA:  
XX  
XX  
Query Match 100.0%; Score 80; DB 14; Length 235;  
Best Local Similarity 100.0%; Pred. No. 2.6e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VIMAGSTNYNSALM 15  
Db 50 viwagstnynsalm 64  
XXXXXXXXXXXXXXXXXXXX  
RESULT 14  
ID AAR32569 standard; Protein; 269 AA.  
XX  
AC AAR32569;  
XX

DT 08-JUN-1993 (first entry)  
XX  
DE Fusion protein encoded by Ox VH-hinge-VL insert.  
XX  
KW Spacer peptide; secretable; single chain; antibody; recombinant;  
KW scab; rDNA; linker; Bos taurus.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FH Region 1..22  
FT /note= "SS"  
FT Region 23..135  
FT /note= "VH"  
FT 136..163  
FT /note= "CBHI hinge"  
FT 164..277  
FT Region /note= "VL"  
FT  
XX  
PN F19103434-A.  
PN WO9302198-A. (First Major Country Equivalent).  
XX  
PD 17-JAN-1992.  
XX  
PE 16-JUL-1991; 91WO-0913434.  
XX  
PR 16-JUL-1990; 90US-0552751.  
XX  
PA (TERE-) TECH RES CENT FINLAND.  
XX  
PI Alfthan K, Knowles JNC, Laukkanen ML, Sizmann D, Takkinen K;  
PI Teeri TT;  
PI  
XX  
DR WPI; 1992-134225/17.  
DR N-PSDB; AAQ36982.  
XX  
PT Prod. of single chain fusion protein, pref. antibody - comprises  
PT transforming host cells, e.g. E. coli with expression constructs  
PT composed of proteins or domains, linked by spacer peptide(s)  
PT  
XX  
PS Example; Fig 4; 56pp; English.  
XX  
XX The sequence is that of the fusion protein encoded by the Ox  
CC VH-CBH1 hinge-VL insert which was used as part of a method for  
CC cloning secretable, biologically active single chain antibodies  
CC (scabs) and other secretable fusion proteins having at least 2  
CC distinct functional proteins or domains.  
XX  
SQ Sequence 269 AA:  
XX  
XX  
Query Match 100.0%; Score 80; DB 13; Length 269;  
Best Local Similarity 100.0%; Pred. No. 3.1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VIMAGSTNYNSALM 15  
Db 72 viwagstnynsalm 86  
XXXXXXXXXXXXXXXXXXXX  
RESULT 15  
ID AAB81972 standard; Protein; 581 AA.  
XX  
AC AAB81972;  
XX  
DT 03-JUL-2001 (first entry)  
XX  
DE Ganglioside GD2 specific antibody related protein smg ID NO: 31.  
DE Ganglioside; GD2; complementation determining region; CDR; antibody;  
KW mouse; cancer.  
XX

OS Synthetic.  
 XX  
 PN WO200123573-A1.  
 XX  
 PD 05-APR-2001.  
 XX  
 PE 29-SEP-2000; 2000WO-JP06773.  
 XX  
 PR 30-SEP-1999; 99JP-0278290.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOSYO KK.  
 XX  
 PI Hanai N, Shitara K, Nakamura K, Niwa R;  
 XX  
 DR WPI; 2001-266163/27.  
 XX  
 PT Human type complementation-determining domain transplanted antibody and  
 PT derivatives against ganglioside GD2, useful in diagnosis and therapy of  
 PT e.g. tumours, has low antigenicity, little side effects but potent  
 PT activity in cancer  
 XX  
 PS Example 3; Page 111-114; 123pp; Japanese.  
 XX  
 CC The present invention describes an antibody, which can react specifically  
 CC with ganglioside GD2, and is transplanted with a human type  
 CC complementation-determining domain (CDR), or its fragments. The antibody  
 CC and its derivatives are useful in diagnosis and therapy of tumours,  
 CC particularly cancer diagnosis. The present sequence is a protein  
 CC used in the exemplification of the invention.  
 XX  
 SQ Sequence 581 AA;

Query Match 100.0%; Score 80; DB 22; Length 581;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
 |||||||||||||  
 Db 50 VIMAGSTNYSALM 64

Search completed: August 20, 2002, 13:15:10  
 Job time: 104 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:16:02 ; Search time 46.28 Seconds  
(Without alignments)  
31.144 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_47\_61  
Perfect score: 80  
Sequence: 1 VIMAGSTNYSALM 15

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	95	2 S17605	Ig heavy chain V r
2	80	100.0	100	2 S14490	Ig heavy chain V r
3	80	100.0	106	2 S26322	Ig heavy chain V r
4	80	100.0	106	2 S14489	Ig heavy chain V r
5	80	100.0	107	2 S14492	Ig heavy chain V r
6	80	100.0	107	2 S14493	Ig heavy chain V r
7	80	100.0	109	2 S11109	Ig heavy chain V r
8	80	100.0	112	2 S11100	Ig heavy chain V r
9	80	100.0	112	2 S11108	Ig heavy chain V r
10	80	100.0	113	2 S11101	Ig heavy chain V r
11	80	100.0	114	2 S11099	Ig heavy chain V r
12	80	100.0	114	2 S11106	Ig heavy chain V r
13	80	100.0	115	2 S11103	Ig heavy chain V r
14	80	100.0	116	2 S11102	Ig heavy chain V r
15	80	100.0	117	2 S10111	Ig heavy chain V r
16	80	100.0	120	2 PL0087	Ig heavy chain V r
17	80	100.0	141	2 S52446	Ig heavy chain V r
18	77	96.2	139	2 A32456	Ig heavy chain pre
19	76	95.0	109	2 PH1025	Ig heavy chain V r
20	76	95.0	114	2 S11104	Ig heavy chain V r
21	76	95.0	114	2 S11105	Ig heavy chain V r
22	75	93.8	112	2 S11098	Ig heavy chain V r
23	71	88.8	107	2 S14491	Ig heavy chain V r
24	70	87.5	118	2 P00266	Ig heavy chain V r
25	69	86.2	140	2 S55028	Ig heavy chain V r
26	67	83.8	116	2 S42484	Ig heavy chain V r
27	67	83.8	118	2 S32786	Ig heavy chain (an
28	66	82.5	231	2 PC4155	Ig gamma-2b chain
29	63	78.8	107	2 S14506	Ig heavy chain V r

30	63	78.8	121	2 S33131	Ig heavy chain V r
31	62	77.5	110	2 PH1024	Ig heavy chain V r
32	61	76.2	90	2 A49042	Ig heavy chain V r
33	61	76.2	100	2 A25913	Ig heavy chain pre
34	61	76.2	115	2 S26470	Ig heavy chain V r
35	61	76.2	115	2 S11107	Ig heavy chain V r
36	61	76.2	116	1 G1MS10	Ig heavy chain pre
37	61	76.2	116	2 A33932	Ig mu chain precu
38	61	76.2	127	2 B31807	Ig heavy chain V r
39	61	76.2	135	2 S31913	Ig gamma-2A chain
40	61	76.2	140	2 S14238	Ig gamma-1 chain p
41	60	75.0	117	2 S38563	Ig heavy chain V r
42	60	75.0	121	2 D30560	Ig heavy chain V r
43	60	75.0	122	2 A49049	Ig heavy chain V r
44	59	73.8	97	2 S55372	Ig heavy chain V r
45	59	73.8	101	2 S03466	Ig heavy chain V r

## ALIGNMENTS

RESULT 1  
S17605  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 06-Jun-1997  
C:Accession: S17605  
R:Jackson, T.; Hoogenboom, H.R.; Griffiths, A.D.; Winter, G.  
Nature 352, 624-628, 1991  
A:Title: Making antibody fragments using phage display libraries.  
A:Reference number: S17230; MUID:91326098  
A:Accession: S17605  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-95 <CLAS>  
C:Superfamily: Immunoglobulin V region: immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:8-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VIMAGSTNYSALM 15  
Db 43 VIMAGSTNYSALM 57

RESULT 2  
S14490  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S14490  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL data library, March 1991  
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in V  
A:Reference number: S14484  
A:Accession: S14490  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-100 <CHED>  
A:Cross-references: EMBL:X58647; NID:951283; PIDN:CAA41504.1; PID:951284  
C:Superfamily: Immunoglobulin V region: immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 4.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
|||||  
Db 50 VIMAGSTNYSALM 64

## RESULT 3

S25322  
Ig heavy chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Mar-1998 #sequence\_revision 19-Mar-1998 #text\_change 21-Jan-2000  
C:Accession: S26322  
R:Stark, S.E.; Caton, A.J.  
J. Exp. Med. 174, 613-624, 1991  
A:Title: Antibodies that are specific for a single amino acid interchange in a protein  
A:Reference number: S26309; MUID:91341421  
A:Accession: S26322  
A:Molecule type: mRNA  
A:Residues: 1-106 <STRA>  
A:Cross-references: EMBL:X59182  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:15-91/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
|||||  
Db 44 VIMAGSTNYSALM 58

## RESULT 4

S14489  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S14489  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL Data Library, March 1991  
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in VH C  
A:Reference number: S14484  
A:Accession: S14489  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-106 <CHE>  
A:Cross-references: EMBL:X58646; NID:951281; PIDN:CAA41503.1; PID:951282  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
|||||  
Db 50 VIMAGSTNYSALM 64

## RESULT 5

S14492  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S14492  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL Data Library, March 1991  
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in VH C  
A:Reference number: S14484  
A:Accession: S14492

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-107 <CHE>  
A:Cross-references: EMBL:X58649; NID:951287; PIDN:CAA41506.1; PID:951288  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 4.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
|||||  
Db 50 VIMAGSTNYSALM 64

## RESULT 6

S14493  
Ig heavy chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 23-Jul-1999  
C:Accession: S14493  
R:Chen, Q.; Stenzel-Poore, M.; Rittenberg, M.B.  
submitted to the EMBL Data Library, March 1991  
A:Description: Natural polypeptide antibodies differ from Ag-induced antibodies in V  
A:Reference number: S14484  
A:Accession: S14493  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-107 <CHE>  
A:Cross-references: EMBL:X58650; NID:951289; PIDN:CAA41507.1; PID:951290  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin  
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 4.4e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
|||||  
Db 50 VIMAGSTNYSALM 64

## RESULT 7

S11109  
Ig heavy chain V region (clone NQ5-89.4) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 21-Jan-2000  
C:Accession: S11109  
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.  
Nature 304, 320-324, 1983  
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazol  
A:Reference number: S07331; MUID:83271467  
A:Accession: S11109  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-109 <KAA>  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
F:15-93/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 109;  
Best Local Similarity 100.0%; Pred. No. 4.5e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYSALM 15  
|||||  
Db 50 VIMAGSTNYSALM 64

```
RESULT 8
S11100
Ig heavy chain V region (clone NQ2-20.5.3) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11100
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11100
A:Molecule type: mRNA
A:Residues: 1-112 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
   |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 9
S11108
Ig heavy chain V region (clone NQ5-78.2.6) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11108
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11108
A:Molecule type: mRNA
A:Residues: 1-112 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
   |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 10
S11101
Ig heavy chain V region (clone NQ2-48.2.2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 21-Jan-2000
C:Accession: S11101
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11101
A:Molecule type: mRNA
A:Residues: 1-113 <KAA>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
```

```
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
   |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 11
S11099
Ig heavy chain V region (clone NQ2-17.4.1) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11099
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11099
A:Molecule type: mRNA
A:Residues: 1-114 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
   |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 12
S11106
Ig heavy chain V region (clone NQ5-96.2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11106
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11106
A:Molecule type: mRNA
A:Residues: 1-114 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 80; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VIMAGSTNYNSALM 15
   |||||||
Db 50 VIMAGSTNYNSALM 64

RESULT 13
S11103
Ig heavy chain V region (clone NQ5-61.1.2) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 23-Aug-1997 #text_change 21-Jan-2000
C:Accession: S11103
R:Kaartinen, M.; Griffiths, G.M.; Markham, A.F.; Milstein, C.
Nature 304, 320-324, 1983
A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone
A:Reference number: S07331; MUID:83271467
A:Accession: S11103
A:Molecule type: mRNA
A:Residues: 1-115 <NAT>
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
```

F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 115;

Best Local Similarity 100.0%; Pred. No. 4.7e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15

|||||

DB 50 VIMAGSTNYNSALM 64

RESULT 14

S11102

Ig heavy chain V region (clone NQ5-4.3.1) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 19-Mar-1997 #sequence\_revision 23-Aug-1997 #text\_change 21-Jan-2000

C:Accession: S11102

R:Kaartinen, M.; Giffiths, G.M.; Markham, A.F.; Milstein, C.

Nature 304, 320-324, 1983

A:Title: mRNA sequences define an unusually restricted IgG response to 2-phenylloxazalone

A:Reference number: S07331; MUID:83271467

A:Accession: S11102

A:Molecule type: mRNA

A:Residues: 1-116 <NAT>

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 116;

Best Local Similarity 100.0%; Pred. No. 4.8e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15

|||||

DB 50 VIMAGSTNYNSALM 64

RESULT 15

S10111

Ig heavy chain V region (clone 26) precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C:Accession: S10111

R:Kaartinen, M.; Solin, M.L.; Maekela, O.

EMBO J. 8, 1743-1748, 1989

A:Title: 'Allelic' forms of immunoglobulin V genes in different strains of mice.

A:Reference number: S10111; MUID:89356648

A:Accession: S10111

A:Molecule type: mRNA

A:Residues: 1-117 <KAA>

C:Cross-references: EMBL:X15471; NID:g50005; PIDN:CAA33499.1; PID:g50006

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: heterotetramer; Immunoglobulin

F:34-116/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 80; DB 2; Length 117;

Best Local Similarity 100.0%; Pred. No. 4.8e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15

|||||

DB 69 VIMAGSTNYNSALM 83

Search completed: August 20, 2002, 13:16:02

Job time: 156 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:51 ; Search time 21.48 Seconds

(without alignments)  
27.039 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_47\_61

Perfect score: 80  
Sequence: 1 VIMAGSTNYNSALM 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	76.2	116	1 HV45_MOUSE	P01821 mus musculus
2	56	70.0	115	1 HV44_MOUSE	P01820 mus musculus
3	55	68.8	507	1 YG46_YEAST	P53301 saccharomyc
4	52	65.0	144	1 HV43_MOUSE	P01819 mus musculus
5	43	53.8	317	1 P1MT_THEMA	O56303 thermotoga
6	40	50.0	311	1 YD17_SCHPO	P87055 schistosach
7	40	50.0	318	1 NU1M_ATEPA	O78693 atelies panl
8	40	50.0	318	1 NU1M_CEBAP	O78695 homo sapien
9	40	50.0	318	1 NU1M_HUMAN	O96126 hylobates 1
10	40	50.0	318	1 NU1M_HYLLA	O92xy4 papi hamad
11	40	50.0	318	1 NU1M_PAPHA	P92278 pongo pygma
12	40	50.0	318	1 NU1M_PONPA	O78694 saguinus oe
13	40	50.0	318	1 NU1M_PONPP	P46596 candida alb
14	40	50.0	318	1 NU1M_SAGOE	O78697 drosophila
15	39	48.8	402	1 OPS4_CANAL	O78707 trichosurus
16	39	48.8	440	1 DCO_DROME	O47868 alligator m
17	38	47.5	318	1 NU1M_TRIUV	O79670 pelomedusa
18	38	47.5	321	1 NU1M_ALIMI	O43151 homo sapien
19	38	47.5	322	1 NU1M_PELSU	P07103 erwinia chr
20	38	47.5	344	1 Y401_HUMAN	O80903 human papil
21	38	47.5	426	1 GUNZ_ERWCH	O01594 allium sati
22	38	47.5	454	1 VE2_HPV37	P76481 escherichia
23	38	47.5	486	1 ALLN_ALLSA	P19450 acetobacter
24	38	47.5	575	1 YF8K_ECOLI	P26676 human parai
25	38	47.5	1319	1 ACSC_ACEXY	P06331 homo sapien
26	38	47.5	2262	1 RRP_PIT2HT	O924n4 klebsiella
27	37	46.9	146	1 HV21_HUMAN	O924n4 carassius a
28	37	46.2	103	1 MCEA_KLEPN	O02988 pleurodeles
29	37	46.2	116	1 HV05_CARAU	P57624 buchiera ap
30	37	46.2	172	1 LECA_PLEMA	P05047 sarcophaga
31	37	46.2	265	1 CYSO_BUCAI	O67886 equilex aeo
32	37	46.2	283	1 LECA_SARPE	
33	37	46.2	453	1 HEMN_AOUAE	

34	37	46.2	475	1 PPB_SERMA	P19147 serratia ma
35	37	46.2	476	1 LA14_LYCES	P29535 lycopersico
36	37	46.2	484	1 DCOR_NEUCR	P27121 neurospora
37	37	46.2	499	1 VG02_BRP22	P26745 bacterioph
38	37	46.2	604	1 AMYC_RH10R	P27683 rhizopus or
39	37	46.2	881	1 NIAL_PHAVU	P39865 phaseolus v
40	36	45.6	116	1 HV60_MOUSE	P18531 mus musculu
41	36	45.0	73	1 MALE_PHOLU	P41130 photorhabdu
42	36	45.0	121	1 HV01_MOUSE	P01745 mus musculu
43	36	45.0	230	1 P1MT_WHEAT	O43209 triticum ae
44	36	45.0	281	1 P0T1_ECOLI	P31136 escherichia
45	36	45.0	301	1 MPT5_MYCLE	O05868 mycobacteri

## ALIGNMENTS

RESULT 1					
HV45_MOUSE		STANDARD;	PRT;	116 AA.	
ID HV45_MOUSE					
AC P01821;					
DT 21-JUL-1986 (Rel. 01, Created)					
DT 21-JUL-1986 (Rel. 01, Last sequence update)					
DT 15-JUL-1999 (Rel. 38, Last annotation update)					
DE Ig heavy chain V region MC101 precursor.					
OS Mus musculus (Mouse).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX NCBI_TaxID=10090;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=82075900; PubMed=6273429;					
RA Kataoka T., Nikaido T., Miyata T., Moriaki K., Honjo T.;					
RT "The nucleotide sequences of rearranged and germline immunoglobulin					
RT VH genes of a mouse myeloma MC101 and evolution of VH genes in					
RL mouse."					
RL J. Biol. Chem. 257:277-285(1982).					
CC					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC					
DR EMBL: J00502; AAA38515.1; -					
DR PIR: A02096; GIMSI0.					
DR InterPro: IPR003006; Ig_MHC.					
DR InterPro: IPR003596; Ig_V.					
DR Pfam: PF00047; Ig; 1.					
DR SMART: SM00406; IGV; 1.					
DR Immunoglobulin V region; Signal.					
KW SIGNAL					
FT CHAIN 1					
FT CHAIN 20					
FT CHAIN 116					
FT NON_TER 116					
SO SEQUENCE 116 AA; 12593 MW; 8079A6E7C552B3E CRC64;					
Query Match					
Best Local Similarity 66.7%; Pred. No. 0.0017;					
Matches 10; Conservative 4; Mismatches 1; Indels 0; Gaps 0;					
QY 1 VIMAGSTNYNSALM 15					
DB 69 VIMSGSTDYNAFT 83					
RESULT 2					
HV44_MOUSE		STANDARD;	PRT;	115 AA.	
ID HV44_MOUSE					
AC P01820;					
DT 21-JUL-1986 (Rel. 01, Created)					

```

DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V region PJ14 precursor.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=81012133; PubMed=6774258;
RA Sakano H., Maki R., Kurosawa Y., Roeder W., Tonegawa S.;
RT "Two types of somatic recombination are necessary for the generation
RT of complete immunoglobulin heavy-chain genes.";
RL Nature 286:676-683(1980).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V00767; CAA24148.1; -.
DR PIR: A02095; HVMS14.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; 1g; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region; Signal.
FT SIGNAL 1 19
FT CHAIN 20 115 IG HEAVY CHAIN V REGION PJ14.
FT NON_TER 115 115
SO SEQUENCE 115 AA; 12447 MW; 7569DD4A4843D500 CRC64;

Query Match
Best Local Similarity 70.0%; Score 56; DB 1; Length 115;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSAL 14
DB 69 MIMGDSSTDYNSAL 82

RESULT 3
YGA6_YEAST STANDARD; PRT; 507 AA.
ID YGA6_YEAST STANDARD; PRT; 507 AA.
AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 52.8 kDa protein in BUB1-HIP1 intergenic region.
GN YGRI89C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).
CC -----
CC -1- SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z72974; CAA97215.1; -.
DR EMBL: X99074; CAA67525.1; -.
DR HSSP: P23904; IAKR.
DR SGD: S0003421; CRH1.
DR InterPro: IPR000757; Glyco_hydro_16.
DR Pfam: PF00722; Glyco_hydro_16; 1.
KW Hypothetical protein.
FT DOMAIN 63 66 POLY-SER.
FT DOMAIN 301 310 POLY-SER.
FT DOMAIN 345 357 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 467 470 POLY-SER.
SO SEQUENCE 507 AA; 52757 MW; 7D7B61F57AE942C CRC64;

Query Match
Best Local Similarity 68.8%; Score 55; DB 1; Length 507;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 WAGGSTNYNSA 13
DB 233 WAGGETNYNSA 243

RESULT 4
HY43_MOUSE STANDARD; PRT; 144 AA.
ID HY43_MOUSE STANDARD; PRT; 144 AA.
AC P01819;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain V region MOPC 141 precursor.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=81012133; PubMed=6774258;
RA Sakano H., Maki R., Kurosawa Y., Roeder W., Tonegawa S.;
RT "Two types of somatic recombination are necessary for the generation
RT of complete immunoglobulin heavy-chain genes.";
RL Nature 286:676-683(1980).
CC -----
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS TRANSLATED FROM A
CC DIFFERENTIATED GENE ISOLATED FROM A MYELOMA THAT SECRETES IGG2B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J00491; AAA8121.1; -.
DR EMBL: V00768; CAA24149.1; -.
DR PIR: A02094; G2MS14.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; 1g; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin V region; Signal.
FT SIGNAL 1 19
FT CHAIN 20 144 IG HEAVY CHAIN V REGION MOPC 141.
FT NON_TER 144 144
SO SEQUENCE 144 AA; 15759 MW; 8E47A7CB3706D30A CRC64;

Query Match
Best Local Similarity 65.0%; Score 52; DB 1; Length 144;

```



Best Local Similarity 69.2%; Pred. No. 0.069;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Oy 2 IWAGSTNYSAL 14  
|||:|||||  
Db 70 IWNGSTDNSTL 82

RESULT 5  
PIMT\_THEME STANDARD: PRT: 317 AA.  
AC Q56308; 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein-L-isospartate O-methyltransferase (EC 2.1.1.77) (Protein-  
beta-aspartate methyltransferase) (PIMT) (Protein L-isospartyl  
methyltransferase) (L-isospartyl protein carboxyl methyltransferase).  
GN PCM OR TM0704.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96134983; PubMed=8550470;  
RA Swanson R.V., Sanna M.G., Simon M.I.;  
RT "Thermostable chemotaxis proteins from the hyperthermophilic  
bacterium Thermotoga maritima."  
RL J. Bacteriol. 178:484-489(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
Heidelberg S.L., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=99003468; PubMed=9784234;  
RA Ichikawa J.K., Clarke S.;  
RT "A highly active protein repair enzyme from an extreme thermophile:  
the L-isospartyl methyltransferase from Thermotoga maritima."  
RL Arch. Biochem. Biophys. 358:222-231(1998).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
RX PubMed=11080641;  
RA Skinner M.M., Puvathingal J.M., Walter R.L., Friedman A.M.;  
RT "Crystal structure of protein isospartyl methyltransferase. A  
catalyst for protein repair."  
RL Structure 8:1189-1201(2000).  
CC -I- FUNCTION: CATALYZES THE METHYL ESTERIFICATION OF L-ISOSPARTYL  
RESIDUES IN PEPTIDES AND PROTEINS THAT RESULT FROM SPONTANEOUS  
DECOMPOSITION OF NORMAL L-ASPARTYL AND L-ASPARAGINYL RESIDUES. IT  
PLAYS A ROLE IN THE REPAIR AND/OR DEGRADATION OF DAMAGED PROTEINS.  
CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + protein L-beta-  
aspartate = S-adenosyl-L-homocysteine + protein L-beta-aspartate  
methyl ester.  
CC -I- SUBUNIT: MONOMER.  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -I- MISCELLANEOUS: EXTREMELY HEAT STABLE, WITH NO LOSS OF ACTIVITY  
AFTER 60 MIN AT 100 DEGREES CELSIUS. ENZYME ACTIVITY IS OBSERVED  
AT TEMPERATURES AS HIGH AS 93 DEGREES CELSIUS WITH AN OPTIMAL  
ACTIVITY OF 164 NMOL/MIN/MG PROTEIN AT 85 DEGREES CELSIUS.  
CC -I- SIMILARITY: BELONGS TO THE L-ISOSPARTYL/D-ASPARTYL PROTEIN  
METHYLTRANSFERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U30501; AAA96385.1; -  
DR EMBL: AE001742; AAD35786.1; -  
DR PDB: 1DL5; 08-DEC-00.  
DR TIGR: TM0704; -  
DR InterPro: IPR000682; PCMT.  
DR InterPro: IPR000051; SAM\_bind.  
DR Pfam: PF01135; PCMT; 1.  
DR PROSITE: PS01279; PCMT; 1.  
KW transferase; Methyltransferase; 3D-structure; Complete proteome.  
SQ SEQUENCE 317 AA; 36400 MW; 2FE6019571ADDF2C CRC64;

Query Match 53.8%; Score 43; DB 1; Length 317;  
Best Local Similarity 53.3%; Pred. No. 5.2;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
Oy 1 VIMAGSTNYSALM 15  
|: |||:|:|  
Db 79 VLEIGGCTGYNAVM 93

RESULT 6  
YD77\_SCHPO STANDARD: PRT: 311 AA.  
ID YD77\_SCHPO  
AC P87055;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Hypothetical 35.1 kDa protein C57A10.07 in chromosome I.  
GN SPAC57A10.07.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: TO YEAST YOR238W.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z94864; CAB08170.1; -  
KM Hypothetical protein; Transmembrane.  
FT TRANSMEM 34 54 POTENTIAL.  
FT TRANSMEM 76 96 POTENTIAL.  
SQ SEQUENCE 311 AA; 35128 MW; 6EE388FDE2DF3C64 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 311;  
Best Local Similarity 75.0%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Oy 2 IWAGSTN 9  
:|:|||||  
Db 87 VWLGGSTN 94

```

RESULT 7
NUM_ATPRA ID NUM_ATPRA STANDARD; PRT: 318 AA.
AC 078693;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MTND1 OR NDI.
OS Ateles paniscus (Black spider monkey).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.
OX NCBI_TaxID=9510;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=98404151; PubMed=9732458;
RA Cao Y., Janke A., Waddell P.J., Westernman M., Takenaka O., Murata S.,
RA Okada N., Pabo S., Hasegawa M.;
RT "Conflict among individual mitochondrial proteins in resolving the
RL phylogeny of eutherian orders."
RL J. Mol. Evol. 47:307-322(1998).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB010971; BAA32096.1; -
DR InterPro; IPR001694; Resp_chain_NADH_DHI.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_NDI_1; 1.
DR PROSITE; PS00668; COMPLEX1_NDI_2; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 318 AA; 35815 MW; 39BCF5A20D415F6 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 318;
Best Local Similarity 40.0%; Pred. No. 17;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 VIMAGGSTNYSALM 15
:::|::|::|:
Db 116 ILMSGMASNSNYALI 130

RESULT 8
NUM_CEBAP ID NUM_CEBAP STANDARD; PRT: 318 AA.
AC 078695;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MTND1 OR NDI.
OS Cebus apella (Brown-capped capuchin).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Cebus.
OX NCBI_TaxID=9515;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=98404151; PubMed=9732458;
RA Cao Y., Janke A., Waddell P.J., Westernman M., Takenaka O., Murata S.,
RA Okada N., Pabo S., Hasegawa M.;
RT "Conflict among individual mitochondrial proteins in resolving the

```

```

RT phylogeny of eutherian orders." ;
RL J. Mol. Evol. 47:307-322(1998).
CC -I CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -I SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation .
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; AB010973; BAA32098.1; -.
DR InterPro; IPR001694; resp_chain_NADH_DH1.
DR Pfam; PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEXI_NDL_1; 1.
DR PROSITE; PS00668; COMPLEXI_NDL_2; 1.
RW Oxidoreductase; NAD: Ubiquinol; Mitochondrion; Transmembrane.
SQ SEQUENCE 318 AA; 35805 MW; DBBAV93FE59AE00A CRC64;

Query Match      50.0%; Score 40; DB 1; Length 318;
Best Local Similarity 40.0%; Pred No. 17;
Matches    6; Conservative    6; Mismatches    3; Indels    0; Gaps    0.

QY          1 VIMAGSGSTNYSALM 15
Db           116 ILMSGMAINSNYALI 130
              ::::::::::::::
RESULT       9
NM1M_HUMAN   STANDARD:             PRT:     318 AA.
AC            P03886; Q37523;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      NADH-ubiquinol oxidoreductase chain 1 [EC 1.6.5.3].
GN      MTND1 OR ND1.
OS      Homo sapiens (Human) .
OC      Mitochondrion.
CC      Mammalia; Euteria; Primates; Catarrhini; Homini; Homo.
CX      NCBI_TaxID=9606;
LN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=81173052; PubMed=7219534;
RA      Anderson S., Bankier A.T., Barrell B.G., de Brunj M.H.L.,
RA      Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
RA      Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
RT      "Sequence and organization of the human mitochondrial genome.";
RL      Nature 290:457-465(1981).
LN      [2]
RN      SEQUENCE FROM N.A., AND VARIANTS ALA-87 AND ALA-168.
RP      TISSUE=Placenta;
RC      MEDLINE=95132634; PubMed=7530363;
RA      Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;
RA      "Recent African origin of modern humans revealed by complete sequences
RT      of hominoid mitochondrial DNAs.";
RL      Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
LN      [3]
RP      SEQUENCE OF 130-318 FROM N.A.
RX      MEDLINE=81170577; PubMed=6260957;
RA      Sanger F., Coulson A.R., Barrell B.G., Smith A.J.H., Roe B.A.;
RT      "Cloning in single-stranded bacteriophage as an aid to rapid DNA
RT      sequencing." ;
RL      J. Mol. Biol. 143:161-178(1980).
LN      [4]
RN      IDENTIFICATION OF PROTEIN.
RP      MEDLINE=85188293; PubMed=3921850;
RX      Chomyn A., Mariotti F., Cleeter M.W.J., Ragan C.I., Matsuno-Yagi A.,
RA      Hatfeldt Y., Doolittle R.F., Altardi G.;

```

RT "Six unidentified reading frames of human mitochondrial DNA encode  
RT components of the respiratory-chain NADH dehydrogenase.";  
RL Nature 314:592-597(1985).  
RN [5]  
RP VARIANT LHON THR-52.  
RX MEDLINE-92026096; PubMed-1928099;  
RA Howell N., Bindoff L.A., McCullough D.A., Kubacka I., Poulton J.,  
RA Mackey D., Taylor L., Turnbull D.M.;  
RT "Leber hereditary optic neuropathy: identification of the same  
RT mitochondrial ND1 mutation in six pedigrees.";  
RL Am. J. Hum. Genet. 49:939-950(1991).  
RN [6]  
RP VARIANT LHON THR-52.  
RX MEDLINE-91241131; PubMed-1674640;  
RA Huoponen K., Vilkkij J., Aula P., Nikoskelainen E.K., Savontaus M.L.;  
RT "A new mtDNA mutation associated with Leber hereditary optic  
RT neuroretinopathy.";  
RL Am. J. Hum. Genet. 48:1147-1153(1991).  
RN [7]  
RP VARIANT LHON PRO-285, AND VARIANT CYS-277.  
RX MEDLINE-91206406; PubMed-2018041;  
RA Howell N., Kubacka I., Xu M., McCullough D.A.;  
RT "Leber hereditary optic neuropathy: involvement of the mitochondrial  
RT ND1 gene and evidence for an intragenic suppressor mutation.";  
RL Am. J. Hum. Genet. 48:935-942(1991).  
RN [8]  
RP VARIANT LHON HIS-304.  
RX MEDLINE-91144615; PubMed-1900003;  
RA Johns D.R., Berman J.;  
RT "Alternative, simultaneous complex I mitochondrial DNA mutations in  
RT Leber's hereditary optic neuropathy.";  
RL Biochem. Biophys. Res. Commun. 174:1324-1330(1991).  
RN [9]  
RP CHARACTERIZATION OF VARIANT LHON THR-52.  
RX MEDLINE-92070510; PubMed-1959619;  
RA Majander A., Huoponen K., Savontaus M.-L., Nikoskelainen E.,  
RA Wikstrom M.;  
RT "Electron transfer properties of NADH:ubiquinone reductase in the  
RT ND1/3460 and the ND4/1178 mutations of the Leber hereditary optic  
RT neuroretinopathy (LHON).";  
RL FEBS Lett. 292:289-292(1991).  
RN [10]  
RP VARIANTS PRO-205; CYS-255 AND PRO-288.  
RX MEDLINE-92098084; PubMed-1757091;  
RA Matzuki S., Noer A.S., Lertit P., Thyagarajan D., Kapasa R.,  
RA Uthapapool P., Byrne E.;  
RT "Normal variants of human mitochondrial DNA and translation products:  
RT the building of a reference data base.";  
RL Hum. Genet. 88:139-145(1991).  
RN [11]  
RP VARIANT LHON HIS-30.  
RX MEDLINE-93038635; PubMed-1417830;  
RA Johns D.R., Neufeld M.J., Park R.D.;  
RT "An ND-6 mitochondrial DNA mutation associated with Leber hereditary  
RT optic neuropathy.";  
RL Biochem. Biophys. Res. Commun. 187:1551-1557(1992).  
RN [12]  
RP VARIANT VAL-31.  
RX MEDLINE-94010883; PubMed-8104867;  
RA Shofner J.M., Brown M.D., Tortoni A., Loti M.T., Cabell M.F.,  
RA Mirra S.S., Beal M.F., Yang C.-C., Gearing M., Salvo R.L.,  
RA Juncos J.L., Hansen L.A., Crain B.J., Fayad M., Reckord C.L.,  
RA Wallace D.C.;  
RT "Mitochondrial DNA variants observed in Alzheimer disease and  
RT Parkinson disease patients.";  
RL Genomics 17:171-184(1993).  
RN [13]  
RP VARIANT MELAS THR-31.  
RX MEDLINE-96303708; PubMed-8723687;  
RA Jaksch M., Hofmann S., Kaufhold P., Obermaier-Kusser B., Zierz S.,  
RA Gerbitz K.-D.;  
RT "A novel combination of mitochondrial tRNA and ND1 gene mutations in  
RT a syndrome with MELAS, cardiomyopathy, and diabetes mellitus.";

RL Hum. Mutat. 7:358-360(1996).  
RN [14]  
RP VARIANT NIDDM THR-4.  
RX MEDLINE-95251666; PubMed-7733935;  
RA Nakagawa Y., Ikegami H., Yamato E., Takekawa K., Fujisawa T.,  
RA Hamada Y., Ueda H., Uchigata Y., Miki T., Kumahara Y.;  
RT "A new mitochondrial DNA mutation associated with  
RT non-insulin-dependent diabetes mellitus.";  
RL Biochem. Biophys. Res. Commun. 209:664-668(1995).  
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -1- DISEASE: DEFECTS IN MTND1 ARE ONE OF THE CAUSES OF LEBER'S  
CC HEREDITARY OPTIC NEUROPATHY (LHON), A MATERNALLY INHERITED DISEASE  
CC RESULTING IN ACUTE BILATERAL BLINDNESS DUE TO RETINAL DEGENERATION  
CC PREDOMINANTLY IN YOUNG MEN. CARDIAC CONDUCTION DEFECTS AND  
CC NEUROLOGICAL DEFECTS HAVE ALSO BEEN DESCRIBED. RESULTING IN OPTIC  
CC NERVE DEGENERATION AND CARDIAC DYSRHYTHMIA.  
CC -1- DISEASE: DEFECTS IN MTND1 ARE ONE OF THE CAUSES OF MELAS SYNDROME,  
CC A DISEASE CHARACTERIZED BY MITOCHONDRIAL ENCEPHALOPATHY, LACTIC  
CC ACIDOSIS, AND STROKE-LIKE EPISODES.  
CC -1- DISEASE: DEFECTS IN MTND1 COULD BE ASSOCIATED WITH ALZHEIMER'S  
CC DISEASE (AD).  
CC -1- DISEASE: DEFECTS IN MTND1 COULD BE ASSOCIATED WITH NON-INSULIN-  
CC DEPENDENT DIABETES MELLITUS (NIDDM).  
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: V00662; CAA24026.1; -;  
DR EMBL: J01415; AAB58943.1; -;  
DR EMBL: D38112; BAA07290.1; -;  
DR EMBL: M10546; AAA65501.1; -;  
DR PIR: A00407; DNHUN1.  
DR MIM: 516000; -;  
DR MIM: 502500; -;  
DR MIM: 535000; -;  
DR MIM: 540000; -;  
DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
DR Pfam: PF00146; NADHdh; 1.  
DR PROSITE: PS00667; COMPLEX1\_ND1\_1; 1.  
DR PROSITE: PS00668; COMPLEX1\_ND1\_2; 1.  
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane;  
KW Leber's hereditary optic neuropathy; Disease mutation; Polymorphism;  
KW Alzheimer's disease; Diabetes.  
FT VARIANT 4 4  
FT A -> T (IN NIDDM).  
FT /FTid=VAR.004747.  
FT Y -> H (IN LHON; SECONDARY MUTATION; DOES  
FT NOT SEEM TO DIRECTLY CAUSE THE DISEASE).  
FT  
FT VARIANT 30 30  
FT /FTid=VAR.004748.  
FT M -> T (IN MELAS).  
FT /FTid=VAR.004749.  
FT M -> V (IN AD).  
FT /FTid=VAR.004750.  
FT A -> T (IN LHON; PRIMARY MUTATION; MEDIUM  
FT SEVERITY; SOME VISION RECOVERY; 80%  
FT REDUCTION IN ROTENONE-SENSITIVE AND  
FT UBIQUINONE-DEPENDENT ELECTRON TRANSFER  
FT ACTIVITY, WHEREAS THE PROXIMAL NADH  
FT DEHYDROGENASE ACTIVITY OF THE COMPLEX IS  
FT UNAFFECTED).  
FT /FTid=VAR.004751.  
FT T -> A.  
FT /FTid=VAR.011346.  
FT T -> A.  
FT /FTid=VAR.011347.  
FT S -> P.  
FT /FTid=VAR.008587.  
FT Y -> C.  
FT VARIANT 255 255

FT	VARIANT	277	277	/FTIG- <u>VAR_008589</u> . Y -> C.
FT	VARIANT	285	285	/FTIG- <u>VAR_004752</u> . L -> P (IN LHON: SECONDARY MUTATION; DOES NOT SEEM TO DIRECTLY CAUSE THE DISEASE).
FT	VARIANT	288	288	/FTIG- <u>VAR_004753</u> . L -> P. /FTIG- <u>VAR_008589</u> . Y -> H (IN LHON: SECONDARY MUTATION; DOES NOT SEEM TO DIRECTLY CAUSE THE DISEASE). /FTIG- <u>VAR_004754</u> .
SO	SEQUENCE	318 AA:	35660 MW:	2F771ACB95002591 CRC64;

Query Match	50.0%;	Score 40;	DB 1;	Length 318;
Best Local Similarity	40.0%;	Pred. No. 17;		
Matches	6; Conservative	6; Mismatches	3; Indels	0; Gaps
Oy	1	YTWAGSTNYNSALM	15	
	::: ::: :: :: :			
Db	116	ILMSCHMASNSNYALI	130	

RESULT	10
ID	NDIM_HYLLA
AC	Q96126;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN	MTMDL OR NDI OR NADH1.
OS	Hylobates lar (Common gibbon).
OG	Mitochondrion.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX	NCBI_TaxID=9580;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ESTER;
RA	Aranson U., Gullberg A., Xu X.;
RT	"A complete mitochondrial DNA molecule of the white-handed gibbon,
RT	Hylobates lar, and comparison among individual mitochondrial genes of
RT	all hominoid genera.";
RL	Hereditas 124:185-189(1996).
CC	-I- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC	-I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X99256; CAA67628.1; -;
DR	InterPro; IPR001694; Resp_chain_NADH_DH1.
DR	Pfam; PF00146; NADHdh; 1.
DR	PROSITE; PS00667; COMPLEX1_NDI_1; 1.
DR	PROSITE; PS00668; COMPLEX1_NDI_2; 1.
DR	Oxidoreductase; NAD; ubiquinone; Mitochondrion; Transmembrane.
DR	SEQUENCE 318 AA; 35814 MW; 3588AAB0FEFE13711 CRC64;

Query Match	50.0%;	Score 40;	DB 1;	Length 318;
Best Local Similarity	40.0%;	Pred. No. 17;		
Matches	6;	Conservative	6;	Mismatches 3; Indels 0; Gaps 0

```
QY      1 VIWAGGSTNYNSALM 15
        ::|:|::| | | |
Db     116 ILWSGWASNSNYALI 130
```

RESULT 11			
ID	NTJM_PAPHA	STANDARD:	PRT: 318 AA.
AC	Q9ZK14;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).		
GN	MTND1 OR ND1 OR NADH1.		
OS	Papio hamadryas (Hamadryas baboon).		
OG	Mitochondrion.		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;		
OC	Cercopitheciinae; Papio.		
OX	NCBI_TaxID=9557;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99065765; PubMed=9847414;		
RA	Arnason U., Gullberg A., Janke A.;		
RT	"Molecular timing of primate divergences as estimated by two		
RL	nonprimate calibration points.";		
RT	J. Mol. Evol. 47:718-727(1998)."		
CC	-1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.		
CC	-1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL: Y18001: CAA76994.1; -		
DR	InterPro: IPR001694; Resp_chain_NADH_DH1.		
DR	Pfam: PF001146; NADHdh; 1.		
DR	PROSITE: PS00667; COMPLEX1_ND1_1; 1.		
DR	PROSITE: PS00668; COMPLEX1_ND1_2; 1.		
DR	Oxidoreductase; NAD: ubiquinone; Mitochondrion; Transmembrane.		
SO	SEQUENCE 318 AA; 35706 MW; 42B440322CFBA418 CRC64;		

Query Match	Score 40;	DB 1;	Length 318;
Best Local Similarity	40.0%;	Pred. No. 17;	
Matches 6;	Conservative	6;	Mismatches 3;
			Indels 0;
			Gaps 0;
Oy	1	VTWAGSGTNYNSALM	15
	::: :	:::	:
Db	116	ILMSGWASNSNYALI	130

RESULT	12
NTJM_PONPA	
ID	NTJM_PONPA
AC	P92690;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	NADH-bi quinone oxidoreductase chain 1 (EC 1.6.5.3).
GN	ND1 OR NADH1.
OS	Pongo pygmaeus abelii (Sumatran orangutan).
OS	Mitochondrion.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.
OX	NCBI_TaxID=9601;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=YN3-312;
RX	MEDLINE=97032590; PubMed=8875856;
RA	Xu X., Arnason U.;
RT	"The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan."

RL J. Mol. Evol. 43:431-437(1996).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X97707; CAA66283.1; -  
 DR InterPro: IPR001694; Resp\_chain\_NADH\_DHL.  
 DR Pfam: PF00146; NADHdh: 1.  
 DR PROSITE: PS00667; COMPLEX1\_ND1\_1: 1.  
 DR PROSITE: PS00668; COMPLEX1\_ND1\_2: 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 SQ SEQUENCE 318 AA; 35512 MW; 94E56A76978EC77A CRC64;

Query Match 50.0%; Score 40; DB 1; Length 318;  
 Best Local Similarity 40.0%; Pred. No. 17;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15  
 DB 116 ILMSGMASNSNYALI 130

RESULT 13  
 NUIM\_PONPP STANDARD; PRT; 318 AA.  
 AC P92718: P92724: 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
 GN MTRND1 OR NDI OR NADH1.  
 OS Pongo pygmaeus pygmaeus (Bornean orangutan).  
 CC Mitochondrion.  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.  
 CC NCBI\_TaxID=9602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISOLATE ANNA, AND ISOLATE DENNIS;  
 RX MEDLINE=97032590; PubMed=8873836;  
 RA Xu X., Arnason U.;  
 RT "The mitochondrial DNA molecule of Sumatran orangutan and a molecular  
 RT proposal for two (Bornean and Sumatran) species of orangutan.";  
 RL J. Mol. Evol. 43:431-437(1996).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X97718; CAA66304.1; -  
 DR EMBL: X97713; CAA66299.1; -  
 DR InterPro: IPR001694; Resp\_chain\_NADH\_DHL.  
 DR Pfam: PF00146; NADHdh: 1.  
 DR PROSITE: PS00667; COMPLEX1\_ND1\_1: 1.  
 DR PROSITE: PS00668; COMPLEX1\_ND1\_2: 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 FT VARIANT 68 I -> T (IN ISOLATE DENNIS).  
 SQ SEQUENCE 318 AA; 35529 MW; F0896A7A33E007ED CRC64;

Query Match 50.0%; Score 40; DB 1; Length 318;  
 Best Local Similarity 40.0%; Pred. No. 17;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15  
 DB 116 ILMSGMASNSNYALI 130

RESULT 14  
 NUIM\_SAGOE STANDARD; PRT; 318 AA.  
 AC O78694:  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
 GN MTRND1 OR NDI.  
 OS Saguinus oedipus (cotton-top tamarin).  
 CC Mitochondrion.  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.  
 CC NCBI\_TaxID=9490;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98404151; PubMed=9732458;  
 RA Okada Y., Janke A., Waddell P.J., Westernman M., Takenaka O., Murata S.,  
 RA Kado N., Paabo S., Hasegawa M.;  
 RT "Conflict among individual mitochondrial proteins in resolving the  
 RT phylogeny of eutherian orders.";  
 RL J. Mol. Evol. 47:307-322(1998).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB010972; BAA32097.1; -  
 DR InterPro: IPR001694; Resp\_chain\_NADH\_DHL.  
 DR Pfam: PF00146; NADHdh: 1.  
 DR PROSITE: PS00667; COMPLEX1\_ND1\_1; FALSE-NEG.  
 DR PROSITE: PS00668; COMPLEX1\_ND1\_2: 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 SQ SEQUENCE 318 AA; 35641 MW; 9E1070B917020CE8 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 318;  
 Best Local Similarity 40.0%; Pred. No. 17;  
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSALM 15  
 DB 116 ILMSGMASNSNYALI 130

RESULT 15  
 OPS4\_CANAL STANDARD; PRT; 402 AA.  
 AC P46596:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Opaque-phase-specific protein Op4 precursor.  
 GN OPS4.  
 OS Candida albicans (Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WO-1;  
 RX MEDLINE=9339284; PubMed=8478072;  
 RA Morrow B., Srikantha T., Anderson J., Soli D.R.;  
 RT "Coordinate regulation of two opaque-phase-specific genes during  
 RL while-opaque switching in Candida albicans."; Infect. Immun. 61:1823-1828(1993).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L10735; -; NOT\_ANNOTATED\_CDS.  
 KW Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 402 OPAQUE-PHASE-SPECIFIC PROTEIN OP4.  
 FT DOMAIN 170 175 POLY-SER.  
 FT DOMAIN 176 185 POLY-ALA.  
 FT DOMAIN 341 349 SER-RICH.  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 402 AA; 41340 MW; 0EEC56D0937E9258 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 402;  
 Best Local Similarity 58.3%; Pred. No. 31;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 IWAGSTNYNSA 13  
 |||||: ||  
 DB 165 IWLGGSSSSSA 176

Search completed: August 20, 2002, 13:17:51  
 Job time: 260 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:23 ; Search time 73.7 Seconds  
(without alignments)  
35.209 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_47\_61

Perfect score: 80  
Sequence: 1 VIMAGSTNYNSALM 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	81.2	482	11	091X92
2	51	63.8	121	11	099NG4
3	46	57.5	937	5	017469
4	45	56.2	449	9	09AY20
5	44	55.0	328	16	092X08
6	44	55.0	613	4	096EX0
7	43	53.8	383	16	092N74
8	43	53.8	497	16	0927U5
9	43	53.8	1390	10	049145
10	42	52.5	145	11	092202
11	42	52.5	453	17	097XP6
12	42	52.5	528	11	09CU84
13	42	52.5	600	3	042795
14	42	52.5	662	3	099012
15	42	52.5	696	16	09CHC2
16	41.3	51.9	597	4	09BU10

17	41.5	51.9	618	4	096AA6	096aa6 homo sapien
18	41	51.2	215	2	09JPN0	09jpn0 neisseria m
19	41	51.2	489	2	P72495	P72495 streptomyce
20	41	51.2	489	2	P72497	P72497 streptomyce
21	40	50.0	302	3	042800	042800 aspergillus
22	40	50.0	318	8	09B1M4	09b1m4 homo sapien
23	40	50.0	318	8	09B1L5	09b1l5 homo sapien
24	40	50.0	318	8	09T9W3	09t9w3 pan troglod
25	40	50.0	318	8	09T9Z0	09t9z0 gorilla gor
26	40	50.0	318	8	09T9X9	09t9x9 pongo pygma
27	40	50.0	318	8	09T9X2	09t9x2 pan paniscu
28	40	50.0	318	8	09B2Y9	09b2y9 homo sapien
29	40	50.0	318	8	09B2X5	09b2x5 homo sapien
30	40	50.0	318	8	09B2X4	09b2x4 homo sapien
31	40	50.0	318	8	09B2V7	09b2v7 homo sapien
32	40	50.0	318	8	09B2U3	09b2u3 homo sapien
33	40	50.0	318	8	09G393	09g393 echinops te
34	40	50.0	318	8	094Z91	094z91 macaca sylv
35	40	50.0	318	8	094P91	094p91 homo sapien
36	40	50.0	319	8	09TEH1	09teh1 cavia porce
37	40	50.0	828	9	056785	056785 staphylococ
38	40	50.0	824	10	0940D1	0940d1 arabidopsis
39	40	50.0	825	10	09SH62	09sh62 arabidopsis
40	40	50.0	1013	10	09LK23	09lk23 arabidopsis
41	40	50.0	1013	10	093Z55	093z55 arabidopsis
42	40	50.0	1021	10	09LHH8	09lhh8 arabidopsis
43	40	50.0	1088	10	09C7C9	09c7c9 arabidopsis
44	40	50.0	1194	2	09R6P0	09r6p0 agrobacteri
45	40	50.0	1451	10	049140	049140 arabidopsis

## ALIGNMENTS

RESULT 1  
091X92 PRELIMINARY: PRT: 482 AA.  
ID 091X92;  
AC 091X92;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE UNKNOWN (PROTEIN FOR MGC:18822).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC011181; AAH11181.1; -  
SQ SEQUENCE 482 AA: 51865 MW: 312E01F9C1BC7F3C CRC64;

Query Match 81.2%; Score 65; DB 11; Length 482;  
Best Local Similarity 85.7%; Pred. NO. 0.0095;  
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 VIMAGSTNYNSAL 14  
Db 69 VIMTGVTNYNSAL 82  
RESULT 2  
099NG4 PRELIMINARY: PRT: 121 AA.  
ID 099NG4;  
AC 099NG4;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SINGLE CHAIN FV (FRAGMENT).  
OS Mus musculus (Mouse).

```

Query Match          63.8%;  Score 51;  DB 11;  length 121;
Best Local Similarity 60.0%;
Matches      9;  Conservative      3;  Mismatches      3;  Indels      0;  Gaps      0

```

RESULT	3	
ID	017469	PRELIMINARY;
	017469	PRT;
	017469	937 AA

RP SEQUENCE FROM N.A.  
RX MEDLINE=97477445: PubMed=9334401:  
RA Colbert H.A., Smith T.L., Bargmann C.I.,  
RT "OSM-9, a novel protein with structural similarity to channels, is  
RT required for olfaction, mechanosensation, and olfactory adaptation in  
RT *Caenorhabditis elegans*,"  
RL J. Neurosci. 17:8259-8269(1997).

RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2.  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode *C. elegans*: a platform for  
investigating biology. The *C. elegans* Sequencing Consortium.";  
LT Science 262:2012-2018(1998).

RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Du Z., Maggi L.;  
RT "The sequence of C. elegans cosmid B0212";  
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ  
DR EMBL; AF031408; AAB87064.1; -;

Query Match	57.58;	Score 46;	DB 5;	Length 937;
Best Local Similarity	90.08;	Pred. No. 32;		
Matches	9;	Conservative	1;	Mismatches 0;
				Indels 0;
				Gaps 0;

RESULT	4	
Q9AYZ0		
ID	Q9AYZ0	PRELIMINARY;
NO	Q9AYZ0	PRT;
		449 AA

DT 01-JUN-2001 (TEMBLrel. 17, Created)  
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
DE DNA TRANSFER PROTEIN.  
GN 20.  
OS Bacteriophage HK620.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=153148;  
RN 111  
RP SEQUENCE FROM N.A.  
RA Clark A.J., Inwood W.B., Cloutier T., Dhillon T.S.;  
RT "Nucleotide Sequence of Coliphage HK620 and the Evolution of Lambdoid  
RT Phages";  
RL J. Mol. Biol. 0:0-0(2001).  
RN 121  
RP SEQUENCE FROM N.A.  
RA Clark A.J., Inwood W.B., Cloutier T., Dhillon T.S.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF353538; AAK28501.1; -  
SQ SEQUENCE 449 AA; 48903 MW; D65283FD2860B7AB CRC64;

Query Match	56.2%	Score 45;	DB 9;	Length 449;
Best Local Similarity	53.8%	Pred. No. 21;		
Matches	7;	Conservative	3;	Mismatches 3; Indels 0; Gaps 0;

RESULT	5	
Q92X08		
ID	Q92X08	PRELIMINARY; PRT; 328 AA

DT 01-DEC-2001 (TREMBLrel.. 19, Created)  
DT 01-DEC-2001 (TREMBLrel.. 19, last sequence update)  
DT 01-DEC-2001 (TREMBLrel.. 19, last annotation update)  
DE PUTATIVE TRANSCRIPTIONAL REGULATOR PROTEIN.  
GN SMB20154.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OG Plasmid pSymB (megaplasmid 2).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.



RC STRAIN-1021;  
 RA MEDLINE-21396508: PubMed-11481431;  
 RA Finan T.M., Weidner S., Mong K., Buhmester J., Chain P., Guzy J.,  
 RA Vorhoeelter F.J., Hernandez-Lucas I., Becker A., Cowie A., Guzy J.,  
 RA "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliloti.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 DR EMBL: AL603642: CAC48554.1: -  
 KW plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 328 AA; 35502 MW; 509F3D1697B9362 CRC64;

Query Match 55.0%; Score 44; DB 16; Length 328;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 VIMAGSTNYNSAL 14  
 ||| | | | | | | | | | | | | | | |  
 DB 214 VAMSGGPTSYESGV 227

RESULT 6  
 O96EYO PRELIMINARY; PRT; 613 AA.  
 AC O96EYO;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE UNKNOWN (PROTEIN FOR MGC:20337).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PRIMARY B-CELLS FROM TONSILS;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC011857; AAH11857.1: -  
 SQ SEQUENCE 613 AA; 67273 MW; 31214203FB8421E7 CRC64;

Query Match 55.0%; Score 44; DB 4; Length 613;  
 Best Local Similarity 61.5%; Pred. No. 43;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 IWAGSTNYNSAL 14  
 ||| | | | | | | | | | | | | | | |  
 DB 70 IYSGSTNYNPSL 82

RESULT 7  
 O92N74 PRELIMINARY; PRT; 383 AA.  
 AC O92N74;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE HYDROLASE/PEPTIDASE PROTEIN (EC 3.).  
 GN SMC02684.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RA MEDLINE-21368234: PubMed-11474104;  
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,  
 RA Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Botte G.,  
 RA Boutry M., Bowser L., Buhmester J., Cadieu E., Capela D., Chain P.,  
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,

RA Gloux S., Godrie T., Goffeau A., Golding B., Guzy J., Gurjal M.,  
 RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,  
 RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Laure V.,  
 RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portelle D., Purnelle B.,  
 RA Ransperger U., Surzycki R., Thebault P., Vandebol M.,  
 RA Vorhoeelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.,  
 RT "The complete genome of the legume symbiont Sinorhizobium meliloti.";  
 RL Science 293:668-672(2001).  
 DR EMBL: AL591790: CAC46925.1: -  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 383 AA; 42517 MW; 1598247DDF63DEBB CRC64;

Query Match 53.8%; Score 43; DB 16; Length 383;  
 Best Local Similarity 53.8%; Pred. No. 37;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 WAGSTNYNSALM 15  
 ||| | | | | | | | | | | | | | | |  
 DB 238 WAGSAYHYHAMM 250

RESULT 8  
 O92705 PRELIMINARY; PRT; 497 AA.  
 AC O92705;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE LIN2693 PROTEIN.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CLIP 11262 / SEROVAR 6A;  
 RX PubMed-11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Blocher H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jodens L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurpark G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,  
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL: AL596173; CAC97919.1: -  
 KW Complete proteome.  
 SQ SEQUENCE 497 AA; 57291 MW; E505C7027CA290F1 CRC64;

Query Match 53.8%; Score 43; DB 16; Length 497;  
 Best Local Similarity 63.6%; Pred. No. 50;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 WAGSTNYNSA 13  
 ||| | | | | | | | | | | | | | | |  
 DB 123 WAGFRANINTA 133

RESULT 9  
 O49145 PRELIMINARY; PRT; 1390 AA.  
 AC O49145;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Arabidopsis arenosa.  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=38785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CARE-4; TRANSPOSON-EVELEKNEVEL RETROTRANSPOSON;  
 RX MEDLINE=98250692; PubMed=9584105;  
 RA Henikoff S., Comai L.;  
 RT "A DNA methyltransferase homolog with a chromodomain exists in  
 RL multiple polymorphic forms in Arabidopsis.";  
 RT Genetics 149:307-318(1998).  
 DR EMBL: AF039376; AAC02672.1; -  
 DR MEROPS: A11.004; -  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00665; Iye; 1.  
 DR Pfam: PF00098; ZF-CCHC; 1.  
 DR SMART: SM00343; ZNF\_C2HC; 1.  
 KW Zinc-finger.  
 FT NON\_TER  
 SQ SEQUENCE 1390 AA; 154000 MW; FC37187A4FAA05CF CRC64;

Query Match 53.8%; Score 43; DB 10; Length 1390;  
 Best Local Similarity 70.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 WAGSSTNTNS 12  
 Db 1298 WAGDTDNYS 1307

RESULT 10  
 ID 092202 PRELIMINARY; PRT; 145 AA.  
 AC 092202;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RIKEN CDNA 5430416A05 GENE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: BC006884; AA006884.1; -  
 SQ SEQUENCE 145 AA; 16031 MW; 492EB74102B475BC CRC64;

Query Match 52.5%; Score 42; DB 11; Length 145;  
 Best Local Similarity 47.6%; Pred. No. 18;  
 Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

OY 3 WAGGS-----TNYSALM 15  
 Db 84 WNGGSNYQANROTSYNSAKM 104

RESULT 11  
 ID 097XP6 PRELIMINARY; PRT; 453 AA.  
 AC 097XP6;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE CYTOSINE PERMEASE (CODB).

GN CODB.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She O., Slingh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweyer M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus p2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL: AB06780; AAK41877.1; -  
 DR InterPro: IPR001248; Transp\_cyl\_pur.  
 KW Complete proteome.  
 SQ SEQUENCE 453 AA; 48816 MW; EF43F53B35471585 CRC64;

Query Match 52.5%; Score 42; DB 17; Length 453;  
 Best Local Similarity 70.0%; Pred. No. 67;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 VIMAGSTNY 10  
 Db 43 VIMAGGALCY 52

RESULT 12  
 ID 09CU84 PRELIMINARY; PRT; 528 AA.  
 AC 09CU84;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE 5430416A05RIK PROTEIN (FRAGMENT).  
 GN 5430416A05RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C57BL/6J; TISSUE=HEAD;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsunk S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK017312; BAB30687.1; -  
 DR MGD: MGI:1918590; 5430416A05RIK.  
 DR InterPro: IPR000687; RI01\_ZK632\_MJ0444.  
 DR InterPro: IPR001245; Tyr-kinase.

DR Pfam; PF01163; RIO1; 1.  
 DR SMART; SM00090; RIO; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; UNKNOWN\_1.  
 FT NOI\_TFR 528 528  
 SO SEQUENCE 528 AA; 60184 MW; 4AED139E966CD122 CRC64;

Query Match 52.5%; Score 42; DB 11; Length 528;  
 Best Local Similarity 47.6%; Pred. No. 79;  
 Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

OY 3 WAGS-----TNNNSALM 15  
 :|||||  
 Db 84 WNGGSNYQANROTSNYSNAKM 104

RESULT 13  
 ID 042795 PRELIMINARY; PRT; 600 AA.  
 AC 042795;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 GN SERINE/THREONINE PROTEIN KINASE.  
 DE LAPK.  
 OS Colletotrichum trifolii.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
 OC mitosporic Phyllachoraceae; Colletotrichum.  
 OX NCBI\_TaxID=5466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yang Z., Dickman M.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF046923; AAC04357.1; -.  
 DR HSP; P05132; ICTP.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR000961; Pkinase\_C.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; Pkinase\_2.  
 DR Pfam; PF00433; Pkinase\_C; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SO SEQUENCE 600 AA; 66893 MW; C336051D3AE5AA6 CRC64;

Query Match 52.5%; Score 42; DB 3; Length 600;  
 Best Local Similarity 66.7%; Pred. No. 92;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 IWAGGSTNY 10  
 :|||||  
 Db 132 LWAGSTNY 140

RESULT 14  
 ID 099012 PRELIMINARY; PRT; 662 AA.  
 AC 099012;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SERINE/THREONINE PROTEIN KINASE.  
 GN PKT1.  
 OS Trichoderma reesei (Hypocrea jecorina).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Hypocreaceae; Hypocrea.  
 OX NCBI\_TaxID=51453;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-QW9414;  
 RX MEDLINE=94357455; PubMed=8076837;  
 RA Morawetz R., Mischak H., Goodnight J., Lendenfeld T., Mushinsky J.F.,  
 RA Kudrick C.P.;  
 RT "A protein kinase-encoding gene, pkl1, from Trichoderma reesei,  
 RT homologous to the yeast YPK1 and YPK2 (YKR2) genes."  
 RL Gene 146:309-310(1994).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; U05811; AAA57318.1; -.  
 DR HSP; P05132; ICTP.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR000961; Pkinase\_C.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; Pkinase\_1.  
 DR Pfam; PF00433; Pkinase\_C; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SO SEQUENCE 662 AA; 73380 MW; A33BDCD93F2C4C23 CRC64;

Query Match 52.5%; Score 42; DB 3; Length 662;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 IWAGGSTNY 10  
 :|||||  
 Db 204 LWAGSTNY 212

RESULT 15  
 ID 09CHC2 PRELIMINARY; PRT; 696 AA.  
 AC 09CHC2;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE UNKNOWN PROTEIN.  
 GN YICA.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IL1403;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403."  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL; AE006314; AAK04908.1; -.  
 KW Complete proteome.  
 SO SEQUENCE 696 AA; 78889 MW; 52AD11ADA2B11EAC CRC64;

Query Match 52.5%; Score 42; DB 16; Length 696;  
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 VIVAGGSTNY 10  
 :|:|:|  
 Db 148 VLVAGSATNY 157

Search completed: August 20, 2002, 13:17:25  
 Job time: 239 sec

---



PR 10-MAY-1996; 96US-0017466.  
XX (BIOJ ) BIOGEN INC.  
XX Benjamin CD, Burkly LC, Hession C, Whitty A;  
XX WPI: 1998-008885/01.  
DR N-PSDB; AAT97441.  
XX  
XX Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
XX Claim 23; Page 81-82; 111pp; English.  
XX  
XX This polypeptide comprises the heavy chain variable region (VH) of  
CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma  
CC deposited as ATCC 12107, and which is specific for the gamma  
CC constant (gc) chain (see AAW31646) of human cytokine receptors. The  
CC invention provides compositions and methods for inhibiting cytokine  
CC signalling using gc chain blocking agents for the treatment of  
CC immunological diseases such as myasthenia gravis, rheumatoid  
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
CC allergy, asthma, parasitic infection, graft vs. host disease or  
CC psoriasis. Preferred gc blocking agents include Mab C9.B8, its Fab  
CC fragment and an antibody having a light chain variable region  
CC CDR selected from those of CP.B8 VH or a heavy chain variable  
CC region CDR selected from those of CP.B8 VL (see AAW31647).  
XX  
XX Sequence 112 AA:

Query Match 100.0%; Score 52; DB 19; Length 112;  
Best Local Similarity 100.0%; Pred. No. 0.0099;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
| | | | | | | | | |  
DB 95 egstvdsmdy 104

RESULT 2  
AAW80726  
ID AAW80726 standard; Protein: 201 AA.  
XX  
XX AAW80726;  
XX  
XX 24-DEC-1998 (first entry)  
XX  
XX S. pneumoniae prolipoprotein diacylglyceryl transferase.  
XX  
XX Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;  
KW virulence; antibody; infection; detection; treatment; hypothetical;  
KW cell wall biosynthetic; external target; minimal gene set protein.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX WO9826072-A1.  
XX  
XX 18-JUN-1998.  
XX  
XX 09-DEC-1997; 97WO-US22578.  
XX  
XX 13-DEC-1996; 96US-0036281.  
XX  
XX (ELIT ) LILLY & CO ELI.  
XX  
XX Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,  
PI Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR;  
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;  
PI Young Bellido ML;  
XX

DR WPI: 1998-348529/30.  
DR N-PSDB; AAV65284.  
XX  
XX Streptococcus pneumoniae nucleic acid sequences - used in DNA chips  
PT for evaluating gene expression, and identification of virulence  
PT genes  
XX  
XX Claim 3; Page 318; 333pp; English.  
XX

XX This sequence represents a S. pneumoniae prolipoprotein diacylglyceryl  
CC transferase. The invention provides DNA sequences (AAV65201 to AAV65304)  
CC from the Streptococcus pneumoniae genome and corresponding protein  
CC sequences (AAW80605 to AAW80728). The protein sequences are classified as  
CC hypothetical, cell wall biosynthetic, external target, or minimal gene  
CC set proteins. A recombinant host containing a vector comprising any of  
CC the above nucleic acids can be used for the recombinant expression of the  
CC proteins. The invention also provides a DNA chip having arrayed on it at  
CC least 15 base pair fragment of any one or more of these DNA sequences.  
CC The DNA chip can be used methods for evaluating gene expression in S.  
CC pneumoniae and for identifying virulence genes in S. pneumoniae.  
CC Antibodies that selectively bind to the above proteins or peptide  
CC fragments can be used to treat S. pneumoniae infection. The antibodies  
CC can also be used to detect S. pneumoniae cells.

XX Sequence 201 AA:

Query Match 73.1%; Score 38; DB 19; Length 201;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10  
| : | | | : | |  
DB 81 gatvdnldy 89

RESULT 3  
AAW38541  
ID AAW38541 standard; Protein: 266 AA.  
XX  
XX AAW38541;  
XX  
XX 06-NOV-1998 (first entry)  
XX  
XX S. pneumoniae prolipoprotein diacylglyceryl transferase.  
XX  
XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
KW immunological response; inoculation; antibody production; inhibitor;  
KW T cell immune response; antimicrobial compound; bacterial adhesion;  
KW extracellular matrix protein; protein-mediated cell invasion; wound;  
KW pathogenesis.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX WO9743303-A1.  
XX  
XX 20-NOV-1997.  
XX  
XX 14-MAY-1997; 97WO-US07950.  
XX  
XX 14-MAY-1996; 96US-0017670.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Stodola RK;  
XX  
XX WPI: 1998-008793/01.  
DR N-PSDB; AAT98597.  
XX  
XX Novel Streptococcus pneumoniae proteins and related DNA - useful for  
PT diagnosing anti-microbial agents for treatment of bacterial

PT infections  
 XX  
 PS Claim 12; Pages 319-320; 483pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein that, based  
 CC on homology with a S. typhimurium protein, is a prolipoprotein  
 CC diacylglycerol transferase, and is encoded by a DNA of the invention.  
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain  
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
 CC invention can be used to identify compounds which interact with and  
 CC inhibit or activate the activity of the proteins. Antagonists can be  
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
 CC immunisation. They can also be used to induce an immunological response  
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
 CC of the encoding nucleic acids in a vector adequate to produce antibody  
 CC and/or T cell immune responses to protect the animal from disease. The  
 CC proteins can also be used to identify antimicrobial compounds which are  
 CC capable of inhibiting their bioactivity. In particular the proteins of  
 CC the invention can be used to prevent adhesion of bacteria to mammalian  
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
 CC block protein-mediated mammalian cell invasion, and to block the normal  
 CC progression of pathogenesis in infections initiated other than by the  
 CC implantation of in-dwelling devices or other surgical techniques.  
 XX  
 SQ Sequence 266 AA;  
 Query Match 73.1%; Score 38; DB 19; Length 266;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GSTVDSMDY 10  
 Db 146 gatvdnldy 154  
 RESULT 4  
 AAW83009 standard; Protein; 266 AA.  
 XX  
 AC AAW83009;  
 XX  
 DT 01-FEB-1999 (first entry)  
 XX  
 DE Streptococcus pneumoniae Lgt protein.  
 XX  
 KW Streptococcus pneumoniae; Lgt; pneumococcal; diagnosis; infection;  
 KW otitis media; conjunctivitis; pneumonia; bacteraemia; meningitis;  
 KW sinusitis; pleural empyema; endocarditis.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN US5840538-A.  
 XX  
 PD 24-NOV-1998.  
 XX  
 PF 06-AUG-1997; 97US-0906753.  
 XX  
 PR 06-AUG-1997; 97US-0906753.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Petit CM;  
 XX  
 DR WPI: 1999-034037/03.  
 DR N-PSDB; AAV69831.  
 XX  
 PT Polynucleotide encoding pneumococcal Lgt polypeptide - used  
 PT therapeutically and in diagnosis  
 XX  
 PS Claim 12; Column 21-24; 14pp; English.  
 CC The present sequence represents a Streptococcus pneumoniae Lgt

CC protein, where the Streptococcus pneumoniae is of the deposited  
 CC strain (NCIMB 40794). A host cell containing a vector comprising the  
 CC nucleic acid sequence encoding the Lgt protein, can be cultured to  
 CC produce the protein, which is a Streptococcus pneumoniae Lgt gene  
 CC product. The nucleic acid sequence encoding the Lgt protein can be  
 CC used biologically, diagnostically, prophylactically, clinically or  
 CC therapeutically. The types of diseases which can be diagnosed and  
 CC treated are: otitis media; conjunctivitis; pneumonia; bacteraemia;  
 CC meningitis; sinusitis; pleural empyema; and endocarditis.  
 XX  
 SQ Sequence 266 AA;  
 Query Match 73.1%; Score 38; DB 20; Length 266;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GSTVDSMDY 10  
 Db 146 gatvdnldy 154  
 RESULT 5  
 AAY81550 standard; Protein; 267 AA.  
 XX  
 AC AAY81550;  
 XX  
 DT 24-MAY-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae type 4 protein sequence #50.  
 XX  
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 KW pneumococcal disease.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200006737-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB02451.  
 XX  
 PR 27-JUL-1998; 98GB-0016337.  
 PR 19-MAR-1999; 99US-0125164.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Gilbert CFG, Hansbro PM;  
 XX  
 DR WPI: 2000-195300/17.  
 XX  
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of  
 PT pneumococcal diseases and for screening agents capable of antagonizing  
 PT or inhibiting expression of the protein  
 XX  
 PS Claim 1; Page 77; 108pp; English.  
 XX  
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences  
 CC isolated from Streptococcus pneumoniae. AAY05407 to AAY05590 represent  
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
 CC The sequences have antibacterial and antiinflammatory properties.  
 CC The protein sequences, and fragments of them, are useful as immunogens  
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in  
 CC diagnostic assays. The proteins and nucleotides can be useful for the  
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
 CC useful for screening an agent capable of antagonising, inhibiting or  
 CC interfering with the function or expression of the proteins in which the  
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
 CC and meningitis. AAY05591 to AAY05614 represent primers used in the  
 CC exemplification of the present invention.

SQ Sequence 267 AA:

Query Match 73.1%; Score 38; DB 21; Length 267;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10  
|:||||:|  
DB 146 gatvdnldy 154

RESULT 6

ABB64149  
ID ABB64149 standard; Protein: 345 AA.

XX ABB64149;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 19239.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08252.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure: SEQ ID NO 19239; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

XX sequences (AB101840-AB116175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pcl\_sequences.

SQ Sequence 345 AA:

Query Match 73.1%; Score 38; DB 22; Length 345;  
Best Local Similarity 70.0%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
| | | | | | | | | |  
DB 31 essvvdslidy 40

RESULT 7

AAG65785  
ID AAG65785 standard; Protein: 1956 AA.

XX AAG65785;

XX 11-FEB-2002 (first entry)

XX Human SNS1 ion channel protein.

XX Ion channel protein; gating; sensory neurone specific protein; SNS;

XX analgesic; antinflammatory; immunosuppressive; human.

XX Homo sapiens.

XX WO200168681-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-GB01108.

XX 14-MAR-2000; 2000GB-0006024.

XX (GLAX ) GLAXO GROUP LTD.

XX Aitken AJ, Chen MX, Gladwell ZM, Powell AJ, Tate SN;

XX WPI; 2001-590038/66.

XX N-PSDB; AA16961.

XX Altered ion channel protein useful for screening modulators that are

XX useful in the manufacture of a medicament for the treatment of pain,

XX inflammation and hypersensitivity, has acquired sensitivity to gating

XX agent -

XX Disclosure: Page 57-58; 64pp; English.

XX The invention provides an altered ion channel protein having acquired

XX sensitivity to a gating agent. The altered ion channel protein can be

XX expressed by standard recombinant methodology. The altered ion channel

XX protein can be derived from a rat, mouse or human unaltered, gating agent

XX insensitive ion channel proteins. The altered protein is useful in

XX screening assay for the identification of modulators capable of

XX modulating an unaltered gating agent insensitive ion channel protein. The

XX modulator identified by the assay is preferably an antagonist of the

XX altered protein, which is useful in the manufacture of a medicament for

XX the treatment of pain, inflammation or hypersensitivity, and for treating

XX a mammalian patient, particularly human. The present sequence represents

XX a human sensory neurone specific (SNS) 1, a sodium ion channel protein.

SQ Sequence 1956 AA:

Query Match 71.2%; Score 37; DB 22; Length 1956;  
Best Local Similarity 77.8%; Pred. No. 2,6e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMD 9  
| | | | | | | | | |

DB 1085 egstvdcld 1093

RESULT 8

AAB61996  
ID AAB61996 standard; Protein: 1956 AA.

XX AAB61996;

XX 14-MAY-2001 (first entry)

XX Human peripheral nerve sodium channel type 3 (hPN3) protein.

XX Peripheral nerve tetrodotoxin-resistant sodium channel alpha subunit;



KM tetrodotoxin; TTX; peripheral nerve sodium channel type 3; PN3; human;  
 KM sodium channel protein; peripheral nervous system; allodynia; neuropathy;  
 KM hyperalgesia; traumatic injury; antidiabetic; neuroprotective; vulnerary;  
 KM analgesic; anti-HIV.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6184349-B1.  
 XX  
 PD 06-FEB-2001.  
 XX  
 PF 15-APR-1997; 97US-0843417.  
 XX  
 PR 11-OCT-1995; 95US-0511828.  
 XX  
 PA (SYNTE ) SYNTEX USA INC.  
 XX  
 PI Herman R, Delgado SG, Fish LM, Sangameswaran L, Rabert DK;  
 XX  
 DR WPI; 2001-202004/20.  
 DR N-PSDB; AAF57016.  
 XX  
 PT New rat and human tetrodotoxin-resistant, voltage-gated sodium channel  
 PT proteins, present in peripheral nerve tissue, useful as a therapeutic  
 PT target for compounds treating peripheral nervous system disorders -  
 XX  
 PS Claim 2; Fig 6A-K; 86pp; English.  
 XX  
 CC The invention provides purified and isolated rat and human peripheral  
 CC nerve tetrodotoxin (TTX)-resistant sodium channel alpha subunit proteins.  
 CC The rat peripheral nerve sodium channel type 3 (PN3) protein or its human  
 CC homologue are useful for the development of antibodies against PN3 which  
 CC are useful in affinity chromatography to purify recombinant sodium  
 CC channel proteins or polypeptides, or as a research tool. The PN3 proteins  
 CC are useful as therapeutic targets for compounds to treat disorders of the  
 CC peripheral nervous system such as allodynia, hyperalgesia, diabetic  
 CC neuropathy, traumatic injury and acquired immunodeficiency syndrome  
 CC (AIDS)-associated neuropathy. The present sequence represents the human  
 CC hPN3 protein.  
 CC  
 SQ Sequence 1956 AA;

Query Match 71.2%; Score 37; DB 22; Length 1956;  
 Best Local Similarity 77.8%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMD 9  
 DB 1085 egstvdclid 1093

RESULT 9  
 ID ABB59854 standard; Protein: 2005 AA.  
 XX  
 AC ABB59854;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 6354.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX

PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL03957.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 6354; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB85737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2005 AA;

Query Match 69.2%; Score 36; DB 22; Length 2005;  
 Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10  
 DB 11 gngidsmdy 19

RESULT 10  
 AAU71866  
 ID AAU71866 standard; Protein: 275 AA.  
 XX  
 AC AAU71866;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE C. glutamicum metabolic pathway protein encoded by gene #1.  
 XX  
 KM Metabolic pathway protein; MP; lysine biosynthesis pathway;  
 KM methionine biosynthesis pathway; large-scale production of fine chemical;  
 KM Corynebacterium diptheriae; diptheria.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200166573-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 22-DEC-2000; 2000WO-IB02035.  
 XX  
 PR 09-MAR-2000; 2000US-187970P.  
 PR 23-JUN-2000; 2000US-0606740.  
 XX  
 PA (BADT ) BASF AG.  
 XX  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;  
 PI Lee H, Hwang B;  
 XX  
 DR WPI; 2001-582269/65.  
 DR N-PSDB; AAS96076.  
 XX



```

RESULT 12
AAC91878
ID AAG91878 standard; Protein: 277 AA.
XX
AC AAG91878;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5632.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
DR N-PSDB: AAH67097.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17: SEQ ID NO: 5632; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 277 AA;

```

```

Query Match 67.3%; Score 35; DB 22; Length 277;
Best Local Similarity 64.3%; Pred. NO. 67;
Matches 9; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

```

```

OY 1 EGSTVDS---MDY 10
   |||||
   |||||
Db 59 egstvdpslwfndy 72

```

```

RESULT 13
AAB70804
ID AAB70804 standard; Protein: 277 AA.
XX
AC AAB70804;
XX

```

```

DT 04-JUN-2001 (first entry)
XX
XX C. glutamicum dapF protein.
DE
XX Overexpression; fermentation; L-lysine; lysine biosynthesis;
KW diaminopimelate epimerase; dapF.
XX
XX Corynebacterium glutamicum.
XX
XX DE19943587-A1.
XX
XX 15-MAR-2001.
XX
XX 11-SEP-1999; 99DE-1043587.
XX
XX 11-SEP-1999; 99DE-1043587.
XX
XX (DEGS ) DEGUSSA-HUELS AG.
XX
XX Kirchner O, Bathe B, Moeckel B, Hartmann M, Kalinowski J;
XX Puhler A, Pfefferle W;
XX
XX WPI: 2001-258936/27.
XX
XX DR N-PSDB: AAF61391.
XX
XX PT Cloned Corynebacterium glutamicum diaminopimelate epimerase (dapF) gene
XX PT useful for producing dapF-overexpressing coryneform bacteria for the
XX PT fermentative production of L-lysine
XX
XX PS Claim 6; Page 13; 18pp; German.
XX
XX
XX This invention describes a novel cloned Corynebacterium glutamicum
XX diaminopimelate epimerase (dapF), EC 5.1.1.7 (I). The invention also
XX describes (a) a polynucleotide that is at least 70% identical to a
XX polynucleotide encoding a polypeptide with a defined sequence of 277
XX amino acids (II) given in the specification; (b) a polynucleotide that is
XX at least 70% identical to a polynucleotide encoding a polypeptide
XX expressed by the dapF gene in PEC-X799A-dapF in C. glutamicum DSM 12968;
XX (c) a polynucleotide encoding a polypeptide comprising an amino acid
XX sequence that is at least 70% identical to (II); (d) a polynucleotide
XX complementary to the polynucleotide of (a), (b) or (c); and (e) a
XX polynucleotide comprising at least 15 consecutive bases of the
XX polynucleotide sequence of (a), (b), (c) or (d). (I) can be used to
XX produce dapF-overexpressing coryneform bacteria for the fermentative
XX production of L-lysine.
XX
XX SQ Sequence 277 AA;

```

```

Query Match 67.3%; Score 35; DB 22; Length 277;
Best Local Similarity 64.3%; Pred. NO. 67;
Matches 9; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

```

```

OY 1 EGSTVDS---MDY 10
   |||||
   |||||
Db 59 egstvdpslwfndy 72

```

```

RESULT 14
AAM93294
ID AAM93294 standard; Protein: 371 AA.
XX
AC AAM93294;
XX

```

```

DT 06-NOV-2001 (first entry)
XX

```

```

DE Human polypeptide, SEQ ID NO: 2787.
XX

```

```

KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX

```

```

OS Homo sapiens.
XX

```

```

PN EP1130094-A2.
XX

```

XX 05-SEP-2001.  
PD  
XX  
PF 07-JUL-2000; 2000EP-0114089.  
XX  
XX 08-JUL-1999; 99JP-0194486.  
PR 11-JAN-2000; 2000JP-0118774.  
PR 02-MAY-2000; 2000JP-0183765.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
DR WPI; 2001-524255/58.  
DR N-PSDB; AAK94214.  
XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -  
XX  
PS Claim 8; SEQ ID NO 2787; 1380bp + sequence listing; English.  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a polypeptide  
CC encoded by a full length human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
CC  
XX  
SQ Sequence 371 AA;

Query Match 67.3%; Score 35; DB 22; Length 371;  
Best Local Similarity 60.0%; Pred. No. 94;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 EGSTVDSMDY 10  
||:|:|  
Db 87 egatldity 96

RESULT 15  
ABG26659  
ID ABG26659 standard; Protein: 473 AA.  
XX  
AC ABG26659;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #26650.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX

PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS90846.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 57018; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
CC  
XX  
SQ Sequence 473 AA;

Query Match 67.3%; Score 35; DB 22; Length 473;  
Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 EGSTVDSMDY 10  
|||:|:|  
Db 145 egsvirslay 154

Search completed: August 20, 2002, 13:15:11  
Job time: 105 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:16:02 ; Search time 46.28 Seconds  
(without alignments)  
20.763 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_95\_104  
Perfect score: 52  
Sequence: 1 EGSTVDSMDY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 263138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	75.0	1078	2 PC4198	peptide synthetase
2	38	73.1	262	2 E95164	prolipo protein dia
3	38	73.1	266	2 D98030	prolipo protein dia
4	37	71.2	329	2 A49868	cathepsin K (EC 3.
5	35	67.3	260	2 T09023	hypothetical prote
6	35	67.3	280	2 T15813	hypothetical prote
7	35	67.3	351	2 C91182	probable fibrillar
8	35	67.3	351	2 G86028	hypothetical prote
9	35	67.3	484	2 T33492	hypothetical prote
10	34	65.4	257	2 D82119	hypothetical prote
11	34	65.4	268	2 F64024	hypothetical prote
12	34	65.4	274	2 F72689	hypothetical prote
13	34	65.4	617	2 AF2897	conserved hypothet
14	34	65.4	641	2 H97672	hypothetical prote
15	34	65.4	822	2 S56823	probable membrane
16	34	65.4	1210	2 H88451	protein ZC135.3 (1
17	33.5	64.4	303	2 B95101	1-phosphofructokin
18	33.5	64.4	303	2 C97969	1-phosphofructokin
19	33	63.5	74	2 A30475	hypothetical prote
20	33	63.5	99	1 OPHUE	acylphosphatase (E
21	33	63.5	134	2 J01403	acidic seminal flu
22	33	63.5	151	1 GGGACR	globin - horn shel
23	33	63.5	162	2 A87755	protein T21E12.3 (
24	33	63.5	245	2 T03088	conserved hypothet
25	33	63.5	269	2 S61555	xyloglucan endo-1,
26	33	63.5	364	2 E83995	leucine dehydrogen
27	33	63.5	400	2 S29686	penicillin-binding
28	33	63.5	400	2 S29687	penicillin-binding
29	33	63.5	400	2 S49089	penicillin-binding

30	33	63.5	401	2 S29688	penicillin-binding
31	33	63.5	518	2 T24783	hypothetical prote
32	33	63.5	536	2 D84325	Htt17 transducer (
33	33	63.5	680	2 T01403	hypothetical prote
34	33	63.5	682	2 J00420	hypothetical prote
35	33	63.5	694	2 H95012	beta-1,3-glucanase
36	33	63.5	702	2 C97884	hypothetical prote
37	33	63.5	785	2 C97884	hypothetical prote
38	33	63.5	1025	2 S34839	collagen alpha 1(V
39	33	63.5	1957	2 S68453	sodium channel pro
40	32	61.5	100	1 A61449	acylphosphatase (E
41	32	61.5	111	2 T37114	hypothetical prote
42	32	61.5	145	2 S75719	myobdopterin bios
43	32	61.5	187	2 A41489	hypothetical prote
44	32	61.5	246	2 A43579	Vmp7 protein homol
45	32	61.5	260	2 E89965	enterotoxin Seo [i

## ALIGNMENTS

RESULT 1  
PC4198  
peptide synthetase (EC 1.14.17.-) - imperfect fungus (Metarhizium anisopliae) (fragme  
C:Species: Metarhizium anisopliae  
C:Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 03-Nov-2000  
C:Accession: PC4198; PC4199  
R:Bailey, A.M.; Kershaw, M.J.; Hunt, B.A.; Paterson, I.C.; Charnley, A.K.; Reynolds,  
Gene 173, 195-197, 1996  
A:Title: Cloning and sequence analysis of an intron-containing domain from a peptide  
A:Reference number: PC4198; MUID:97082966  
A:Accession: PC4198  
A:Molecule type: DNA  
A:Residues: 1-1078 <BAI>  
A:Cross-references: EMBL:X89442; NID:q1015536; PID:q1015537  
A:Accession: PC4199  
A:Molecule type: Protein  
A:Residues: 58-67,136-146 <BI2>  
C:Genetics:  
A:Gene: pes  
A:Introns: 386/3; 762/2  
C:Superfamily: acyl carrier protein homology; acetate--CoA ligase homology  
C:Keywords: carrier protein; oxidoreductase  
F:32-499/Domain: acetate--CoA ligase homology <ACI>  
F:526-594/Domain: acyl carrier protein homology <ACP3>

Query Match 75.0%; Score 39; DB 2; Length 1078;  
Best Local Similarity 77.8%; Pred. No. 21;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9  
DB 438 DGSIVDSLD 446

RESULT 2  
E95164  
prolipo protein diacylglycerol transferase [imported] - Streptococcus pneumoniae (stra  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: E95164  
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf  
son, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: E95164  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75510.1; PID:g14972901; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPL412  
C:Superfamily: prolipoprotein diacylglycerol transferase

Query Match 73.1%; Score 38; DB 2; Length 262;  
Best Local Similarity 66.7%; Pred. No. 7.7;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GSTVDSMDY 10  
1:||||:|  
DB 142 GATVDMLD 150

RESULT 3  
D98030  
prolipoprotein diacylglycerol transferase (EC 2.4.99.-) [Imported] - Streptococcus pneum  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
C:Accession: D98030  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; H  
e, R.; Leblanc, D.J.; Lee, L.N.; Letkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: D98030  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-266 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAL00073.1; PID:g15458909; GSPDB:GN00174  
C:Genetics:  
A:Gene: lgt  
C:Superfamily: prolipoprotein diacylglycerol transferase  
C:Keywords: glycosyltransferase

Query Match 73.1%; Score 38; DB 2; Length 266;  
Best Local Similarity 66.7%; Pred. No. 7.8;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GSTVDSMDY 10  
1:||||:|  
DB 146 GATVDMLD 154

RESULT 4  
AA9868  
cathpsin K (EC 3.4.22.-) precursor [similarity] - rabbit  
N:Alternate names: osteoclast cysteine proteinase OC-2  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 30-Jun-1995 #sequence\_revision 30-Jun-1995 #text\_change 20-Jun-2000  
C:Accession: AA9868  
R:Tezuka, K.; Tezuka, Y.; Maejima, A.; Sato, T.; Nemoto, K.; Kamiooka, H.; Hakeda, Y.; Ku  
J. Biol. Chem. 269, 1106-1109, 1994  
A:Title: Molecular cloning of a possible cysteine proteinase predominantly expressed in  
A:Reference number: AA9868; MUID:94117413  
A:Accession: AA9868  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-329 <TEZ>  
A:Cross-references: GB:D14036; NID:g454186; PIDN:BA003125.1; PID:g454187  
C:Superfamily: papain  
C:Keywords: cysteine proteinase; hydrolase  
F:139,276,296/active site: Cys, His, Asn #status predicted

Query Match 71.2%; Score 37; DB 2; Length 329;  
Best Local Similarity 70.0%; Pred. No. 15;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10  
1:|11:|1  
DB 112 EGRTPSIDY 121

RESULT 5  
T09023  
hypothetical protein T27E11.80 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
C:Accession: T09023  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16533  
A:Accession: T09023  
A:Molecule type: DNA  
A:Residues: 1-260 <BEV>  
A:Cross-references: EMBL:AL078579; GSPDB:GN00062; ATSP:T27E11.80  
A:Experimental source: cultivar Columbia; BAC clone T27E11  
C:Genetics:  
A:Gene: ATSP:T27E11.80  
A:Map position: 4

Query Match 67.3%; Score 35; DB 2; Length 260;  
Best Local Similarity 66.7%; Pred. No. 30;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9  
:||||11:1  
DB 98 DGSTADSLD 106

RESULT 6  
T15813  
hypothetical protein C46H3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15813  
R:Favell, T.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C46H3.  
A:Reference number: Z16409  
A:Accession: T15813  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-280 <FAV>  
A:Cross-references: EMBL:U41271; NID:g1086855; PID:g1086858; PIDN:AAA82445.1; CESP:C4  
C:Genetics:  
A:Gene: CESP:C46H3.3  
A:introns: 7/3; 59/1; 107/1; 155/3; 187/3; 224/3; 263/3

Query Match 67.3%; Score 35; DB 2; Length 280;  
Best Local Similarity 66.7%; Pred. No. 33;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10  
1:||||:|  
DB 130 GANVDSVDY 138

RESULT 7  
C91182  
probable fibrin protein precursor [Imported] - Escherichia coli (strain O157:H7, su  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C91182  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91182  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-351 <HAV>  
A:Cross-references: GB:BA000007; PIDN:BA837850.1; PID:q13363901; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs4427

Query Match 67.3%; Score 35; DB 2; Length 351;  
Best Local Similarity 77.8%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10  
| | | | | | | | | |  
DB 99 GRTVDSMKY 107

RESULT 8  
G86028  
Probable fimbrial protein Z4966 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G86028

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001.  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G86028  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-351 <STO>  
A:Cross-references: GB:AE005174; NID:q12518274; PIDN:AAG58691.1; GSPDB:GN00145; UWGP:Z49  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z4966

Query Match 67.3%; Score 35; DB 2; Length 351;  
Best Local Similarity 77.8%; Pred. No. 41;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10  
| | | | | | | | | |  
DB 99 GRTVDSMKY 107

RESULT 9  
T33492  
Hypothetical protein C36C9.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T33492

R:Fulton, B.; Martin, J.; O'Brien, D.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of C. elegans cosmid C36C9.  
A:Reference number: Z21357

A:Accession: T33492  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-484 <FUI>  
A:Cross-references: EMBL:AF098986; PIDN:AAC67427.1; GSPDB:GN00028; CESP:C36C9.5  
A:Experimental source: strain Bristol N2; clone C36C9  
C:Genetics:  
A:Gene: CESP:C36C9.5

A:Map position: X  
A:Introns: 31/1; 136/1; 194/1; 257/1; 282/1; 313/1; 344/1; 373/1; 408/1; 437/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C36C9.5

Query Match 67.3%; Score 35; DB 2; Length 484;  
Best Local Similarity 60.0%; Pred. No. 58;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
| | | | | | | | | |  
DB 235 EGSSSETMDY 244

RESULT 10

DB2119  
Hypothetical protein VC2094 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: DB2119  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bess, S.; Qin, H.; Dragol, I.; Sellers  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833

A:Accession: DB2119  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-257 <HEI>  
A:Cross-references: GB:AE004283; GB:AE003852; NID:g9656636; PIDN:AAF95240.1; GSPDB:GN  
A:Experimental source: serogroup O1, strain N16961, biotype El Tor  
C:Genetics:  
A:Gene: VC2094  
A:Map position: 1

Query Match 65.4%; Score 34; DB 2; Length 257;  
Best Local Similarity 60.0%; Pred. No. 47;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
| | | | | | | | | |  
DB 84 EGKTVGSIDF 93

RESULT 11

F64024  
Hypothetical protein H1273 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999

C:Accession: F64024  
R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weisman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630

A:Accession: F64024  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-268 <TIG>

A:Cross-references: GB:U32807; GB:LA42023; NID:q1574725; PIDN:AAC22921.1; PID:q1574728

Query Match 65.4%; Score 34; DB 2; Length 268;  
Best Local Similarity 60.0%; Pred. No. 49;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
| | | | | | | | | |  
DB 80 OGSTVVALDY 89

RESULT 12

F72689  
Hypothetical protein APE0934 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C:Accession: F72689  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamaya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
A:Reference number: A72450; MUID:99310339  
A:Accession: F72689  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <KAN>  
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA7918.1; PID:d1043704; PID:g510  
A:Experimental source: strain KI  
C:Genetics:  
A:Gene: APE0934  
C:Superfamily: Aeropyrum pernix hypothetical protein APE0934

Query Match 65.4%; Score 34; DB 2; Length 274;  
Best Local Similarity 60.0%; Pred. No. 50;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
|||||:  
Db 241 EGSTDSIPF 250

RESULT 13  
AF2897  
conserved hypothetical protein Atu2615 [imported] - Agrobacterium tumefaciens (strain C5  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AF2897  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
: Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AF2897  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-617 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AAL43596.1; PID:g17741114; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu2615  
A:Map position: circular chromosome

Query Match 65.4%; Score 34; DB 2; Length 617;  
Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
|||||:  
Db 336 EASSVEGMDY 345

RESULT 14  
H97672  
hypothetical protein AGR\_C\_4742 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: H97672  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194

A:Accession: H97672  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-641 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK88337.1; PID:g15157817; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_4742  
A:Map position: circular chromosome

Query Match 65.4%; Score 34; DB 2; Length 641;  
Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
|||||:  
Db 360 EASSVEGMDY 369

RESULT 15  
S56823  
probable membrane protein YJL051w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein J1156  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Jun-2000  
C:Accession: S56823  
R:Pohl, T.M.; Aljinovic, G.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56793  
A:Accession: S56823  
A:Molecule type: DNA  
A:Residues: 1-822 <TOV>  
A:Cross-references: EMBL:Z49326; NID:g1008186; PID:g1008187; GSPDB:GN00010; MIPS:YJL0  
C:Genetics:  
A:Gene: MIPS:YJL051w  
A:Map position: 10L  
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YJL051w  
C:Keywords: transmembrane protein

Query Match 65.4%; Score 34; DB 2; Length 822;  
Best Local Similarity 60.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10  
|||||:  
Db 670 EGSTDPINY 679

Search completed: August 20, 2002, 13:16:03  
Job time: 157 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:51 ; Search time 21.48 Seconds

(without alignments)  
18.026 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_95\_104  
Perfect score: 52  
Sequence: 1 EGSTVDSMDY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	37	71.2	1 CATK_RABIT	P43236 oryctolagus
2	37	71.2	1 CATK_PIG	Q091e3 sus scrofa
3	34	65.4	1 YC73_HAEIN	P44150 haemophilus
4	34	65.4	1 YJF1_YEAST	P47046 saccharomyc
5	33	63.5	1 ACYO_HUMAN	P07311 homo sapien
6	33	63.5	1 ASFP_BOVIN	P29392 bos taurus
7	33	63.5	1 GUB_CERRH	P02215 cerithidea
8	33	63.5	1 MER5_ARATH	P24806 arabidopsis
9	33	63.5	1 DHLE_BACCE	Q59194 bacillus ce
10	33	63.5	1 ECEB2_HALEL	O52250 halomonas e
11	33	63.5	1 YRYA_CAEEL	O27519 caenorhabdi
12	33	63.5	1 E13B_BACCT	P23903 bacillus ci
13	33	63.5	1 PDC2_CANAL	O60035 candida alb
14	33	63.5	1 CA16_MOUSE	Q04857 mus musculu
15	32	61.5	1 ACYO_MOUSE	P56376 mus musculu
16	32	61.5	1 ACYO_BOVIN	P41500 bos taurus
17	32	61.5	1 ACYO_PIG	P24540 sus scrofa
18	32	61.5	1 Y903_SYNT3	Q53371 synecocyst
19	32	61.5	1 CATK_HUMAN	P43235 homo sapien
20	32	61.5	1 CATK_MACRA	O77641 macaca fasc
21	32	61.5	1 CATK_MOUSE	P55097 mus musculu
22	32	61.5	1 CATK_RAT	O35186 rattus norv
23	32	61.5	1 Y548_CAEEL	O09374 caenorhabdi
24	32	61.5	1 ABR4_ASPMG	P42254 aspergillus
25	32	61.5	1 682_1	P07596 lactococcu
26	32	61.5	1 YK09_YEAST	P36129 saccharomyc
27	31	59.6	1 GRE4_MYCPN	P78019 mycoplasma
28	31	59.6	1 GRE4_MYCCE	P47524 mycoplasma
29	31	59.6	1 JAG_BACHD	O9rcab bacillus ha
30	31	59.6	1 YXK_CAEEL	O20728 caenorhabdi
31	31	59.6	1 YFC8_YEAST	P43569 saccharomyc
32	31	59.6	1 MCH_ARCFU	O28344 archaeoglob
33	31	59.6	1 YR62_CAEEL	O20347 caenorhabdi

34	31	59.6	397	1	TIG2_YEAST	O08144 saccharomyc
35	31	59.6	417	1	GAB3_YEAST	O04739 saccharomyc
36	31	59.6	428	1	LE22_METTH	O27668 methanobact
37	31	59.6	458	1	YC99_YEAST	P25594 saccharomyc
38	31	59.6	472	1	C15Y_DAUCA	O08433 daucus caro
39	31	59.6	486	1	MORE_BACHD	O9K954 bacillus ha
40	31	59.6	509	1	HMD1_ORYSA	P48019 oryza sativ
41	31	59.6	523	1	SYN_TREPA	O83618 treponema p
42	31	59.6	582	1	YK85_YEAST	P36172 saccharomyc
43	31	59.6	769	1	ITB8_HUMAN	P26012 homo sapien
44	31	59.6	878	1	SYV_METJA	Q58413 methanococc
45	31	59.6	966	1	ST10_MOUSE	O55098 mus musculu

## ALIGNMENTS

RESULT	ID	CATK_RABIT	STANDARD	PRT	329 AA.
AC	P43236:				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Cathepsin K precursor (EC 3.4.22.38) (OC-2 protein).				
GN	CTSK.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
OX	NCBI_Taxid=9986;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Bone;				
RX	MEDLINE=94117413; PubMed=8288568;				
RA	Tezuka K.-I., Tezuka Y., Maejima A., Sato T., Nemoto K.,				
RA	Kamiooka H., Hakeda Y., Kamegawa M.;				
RT	"Molecular cloning of a possible cysteine proteinase predominantly				
RT	expressed in osteoclasts.";				
RL	J. Biol. Chem. 269:1106-1109(1994).				
CC	- FUNCTION: CLOSELY INVOLVED IN OSTEOCLASTIC BONE RESORPTION AND MAY PARTICIPATE PARTIALLY IN THE DISORDER OF BONE REMODELING. DISPLAYS POTENT ENDOPEPTIDASE ACTIVITY AGAINST FIBRINOGEN AT ACID PH. MAY PLAY AN IMPORTANT ROLE IN EXTRACELLULAR MATRIX DEGRADATION.				
CC	- CATALYTIC ACTIVITY: Broad proteolytic activity. With small-molecule substrates and inhibitors, the major determinant of specificity is P2, which is preferably Leu, Met > Phe, and not Arg.				
CC	- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN OSTEOCLASTS (BONES).				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE PAPAEN FAMILY OF THIOLE PROTEASES.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: D14036; BAA03125.1; -				
DR	HSSP: P43235; IMEM.				
DR	MEROPS: C01.036; -				
DR	InterPro: IPR000668; Peptidase.C1.				
DR	InterPro: IPR000169; Thiolprol_act.site.				
DR	Pfam: PF00112; Peptidase.C1; 1.				
DR	PRINTS: PR00705; PAPAEN.				
DR	PROSITE: PS00139; THIOLE PROTEASE CYS; 1.				
DR	PROSITE: PS00639; THIOLE PROTEASE HIS; 1.				
DR	PROSITE: PS00640; THIOLE PROTEASE ASN; 1.				
KW	Hydrolase; Thiol protease; Zymogen; Signal.				
FT	SIGNAL 1 15				
FT	PROPEP 16 114				
FT	CHAIN 115 329				
FT	CATHEPSIN K.				

```

FT ACT_SITE 139 139 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT ACT_SITE 296 296 BY SIMILARITY.
FT DISULFID 136 177 BY SIMILARITY.
FT DISULFID 170 210 BY SIMILARITY.
FT DISULFID 269 318 BY SIMILARITY.
FT CAROHND 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHND 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 329 AA: 36870 MW: 875D8582876B51F CRC64;

Query Match
Best Local Similarity 71.2%; Score 37; DB 1; Length 329;
Best Local Similarity 70.0%; Pred. No. 4.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EGSTVDSMDY 10
Db 112 EGRTPSIDY 121

RESULT 2
CATK_PIG STANDARD; PRT: 330 AA.
AC 09GLE3:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cathepsin K precursor (EC 3.4.22.38).
GN CTSK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP TISSUE=Thyroid;
RC MEDLINE=20534869; PubMed=11082042;
RA Tepel C., Broemme D., Herzog V., Brix K.;
RA "Cathepsin K in thyroid epithelial cells: sequence, localization and
RL possible function in extracellular proteolysis of thyroglobulin.";
RL J. Cell Sci. 113:4487-4498(2000).
CC -1- FUNCTION: Closely involved in the disorder of bone remodeling. Displays
CC participate partially in the disorder of bone remodeling. Displays
CC potent endoprotease activity against fibrinogen at acid pH. May
CC play an important role in extracellular matrix degradation (by
CC similarity). Mediates thyroxine liberation by limited proteolysis
CC of thyroglobulin at neutral pH in vitro.
CC -1- CATALYTIC ACTIVITY: Broad proteolytic activity. With small-
CC molecule substrates and inhibitors, the major determinant of
CC specificity is P2, which is preferably Leu, Met > Phe, and not
CC Arg.
CC -1- SUBCELLULAR LOCATION: Lysosomal and extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAIN FAMILY OF THIOL PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF292030; AAC12340.1; -.
CC InterPro: IPR000668; Peptidase_C1.
CC InterPro: IPR000169; Thiolprol_act_site.
CC Pfam: PF00112; Peptidase_C1; 1.
CC PRINTS: PR00705; PAPAIN
CC PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
CC PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
CC PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
CC Hydrolyase: Thiol protease; Lysosome; Signal; Lysosome; Glycoprotein.
FT SIGNAL 1 16 POTENTIAL.

```

```

FT PROPEP 17 115 ACTIVATION PEPTIDE.
FT CHAIN 116 330 CATHEPSIN K.
FT ACT_SITE 140 140 BY SIMILARITY.
FT ACT_SITE 277 277 BY SIMILARITY.
FT ACT_SITE 297 297 BY SIMILARITY.
FT DISULFID 137 178 BY SIMILARITY.
FT DISULFID 171 211 BY SIMILARITY.
FT DISULFID 270 319 BY SIMILARITY.
FT CAROHND 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 330 AA: 37069 MW: ABE089EFE5F8170 CRC64;

Query Match
Best Local Similarity 71.2%; Score 37; DB 1; Length 330;
Best Local Similarity 70.0%; Pred. No. 4.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EGSTVDSMDY 10
Db 113 EGRTPSIDY 122

RESULT 3
YC73_HAEIN STANDARD; PRT: 268 AA.
AC P44150;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein H11273.
GN H11273
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U32807; AAC22921.1; -.
CC TIGR: H11273;
CC InterPro: IPR000051; SAM_bind.
CC Hypothetical protein; Complete proteome.
KW SEQUENCE 268 AA: 30510 MW: E5B28DA7AAD4D0B CRC64;

Query Match
Best Local Similarity 65.4%; Score 34; DB 1; Length 268;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 EGSTVDSMDY 10
Db 80 QGSTVALDY 89

```

```

RESULT 4
ID YJFL_YEAST STANDARD: PRT: 822 AA.
AC P47046.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DR Hypothetical 91.3 kDa protein in TDPH-MTR4 intergenic region.
GN YJL051W OR J1156.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; Z49326; CAAB9342.1; -.
DR SGD; S0003587; YJL051W.
DR Hypothetical protein; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 121 141 POTENTIAL.
SQ SEQUENCE 822 AA; 91256 MW; BB850C5B409F8820 CRC64;

Query Match
Best Local Similarity 65.4%; Score 34; DB 1; Length 822;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSYVDSMDY 10
Db 670 EGSYVDSMDY 679

RESULT 5
ID ACYO_HUMAN STANDARD: PRT: 98 AA.
AC P07311.
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acylphosphatase, organ-common type isozyme (EC 3.6.1.7)
DE (Acylphosphate phosphohydrolase) (Acylphosphatase, erythrocyte
DE isozyme)
GN ACYPL OR ACYPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RA MEDLINE=87101109; PubMed=3026468;
RA Liguri G., Camici G., Manso G., Cappugi G., Naasi P., Modesti A.,
RA Rampori G.;
RT "A new acylphosphatase isoenzyme from human erythrocytes:
RT purification, characterization, and primary structure.";
RL Biochemistry 25:8089-8094(1986).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Placenta;
RA Raugel G.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Dors M., Dickhoff R., James R.,
RA Loretz C., Lasky S., Madan A., Prescott S., Ratcliffe A., Shaffer T.,
RA Hood L.;
RT "Sequencing of human chromosome 14q24.3 region.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 9-85 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95317414; PubMed=7796909;
RA Flaschi T., Raugel G., Marzocchi R., Chiarugi P., Cirri P.,
RA Rampori G.;
RT Cloning and expression of the cDNA coding for the erythrocyte
RT isoenzyme of human acylphosphatase.";
RL FEBS Lett. 367:145-148(1995).
CC -1- FUNCTION: ITS PHYSIOLOGICAL ROLE IS NOT YET CLEAR.
CC -1- CATALYTIC ACTIVITY: An acyl phosphate + H(2)O = a fatty acid anion
CC + phosphate
CC -1- TISSUE SPECIFICITY: ORGAN-COMMON TYPE ISOZYME IS FOUND IN MANY
CC DIFFERENT TISSUES.
CC -1- SIMILARITY: BELONGS TO THE ACYLPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X84194; CA58987.1; -.
DR EMBL; AC007055; AAD31937.1; -.
DR PIR; A25587; QPH0E.
DR HSSP; P41500; ZACY.
DR MIM; 600875; -.
DR InterPro: IPR001792; Acylphosphatase.
DR Pfam: PF00708; Acylphosphatase; 1.
DR PRINTS; PR00112; ACYLPHPTASE.
DR ProDom; PD001884; Acylphosphatase; 1.
DR PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
DR PROSITE; PS00151; ACYLPHOSPHATASE_2; 1.
KM Hydrolase; Acetylation; Multigene family.
FT INIT_MET 0 0
FT MOD_RES 1 1 ACETYLATION.
FT ACT_SITE 23 23 POTENTIAL.
FT ACT_SITE 41 41 POTENTIAL.
FT CONFLICT 19 19 G -> R (IN REF. 4)
SQ SEQUENCE 98 AA; 11130 MW; 6C5EE71D00258B06 CRC64;

Query Match
Best Local Similarity 63.5%; Score 33; DB 1; Length 98;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSYVDSMDY 10
Db 2 EGNLTISVDY 11

RESULT 6
ID ASFP_BOVIN STANDARD: PRT: 134 AA.
AC P29392.
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Acidic seminal fluid protein precursor (ASFP).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OK NCBI\_TaxID=9913;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Seminal vesicle;  
 RX MEDLINE=92181448; PubMed=1543494;  
 RA Wempe F., Einspanier R., Schelt K.H.;  
 RT "Characterization by cDNA cloning of the mRNA of a new growth factor  
 from bovine seminal plasma: acidic seminal fluid protein.";  
 RL Biochem. Biophys. Res. Commun. 183:232-237(1992).  
 RN (2)  
 RP SEQUENCE OF 21-43.  
 RC TISSUE=Seminal vesicle;  
 RX MEDLINE=91378963; PubMed=1698381;  
 RA Einspanier R., Einspanier A., Wempe F., Schelt K.H.;  
 RT "Characterization of a new bioactive protein from bovine seminal  
 fluid.";  
 RL Biochem. Biophys. Res. Commun. 179:1006-1010(1991).  
 RN (3)  
 RP DISULFIDE BONDS.  
 RX MEDLINE=94237283; PubMed=8181566;  
 RA Einspanier R., Krause I., Calvete J.J., Toepfer-Petersen E.,  
 RT Klostermeyer H., Karg H.;  
 RL "Bovine seminal plasma asf: localization of disulfide bridges and  
 detection of three different isoelectric forms.";  
 FEBS Lett. 344:61-64(1994).  
 RN (4)  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=97475216; PubMed=9334740;  
 RA Romero A., Romao M.J., Varela P.F., Koellin I., Dias J.M.,  
 RA Carvalho A.L., Sanz L., Toepfer-Petersen E., Calvete J.J.;  
 RT "The crystal structures of two spermadhesins reveal the CUB domain  
 fold.";  
 RL Nat. Struct. Biol. 4:783-788(1997).  
 CC -1- FUNCTION: STIMULATES CELL DIVISION AND PROGESTERONE SECRETION  
 OF BOVINE GRANULOSA CELLS IN VITRO IN A POTENT AND DOSE DEPENDENT  
 MANNER. THIS PROTEIN APPEARS TO BE A POTENT GROWTH FACTOR WITH  
 EFFECTS ON OVARIAN GRANULOSA CELLS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: SEMINAL VESICLE TISSUE, AMPULLA AND WEAKLY  
 IN TISSUE OF EPIDIDYMIS.  
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SPERMADHESIN FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: M84603; AAA30745.1; .  
 DR PIR: PH0213; PH0213.  
 DR PIR: J01403; J01403.  
 DR PDB: 1SFP; 24-JUN-98.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR000124; Spermadhesin.  
 DR Pfam: PF00431; CUB; 1.  
 DR SMART: SM00042; CUB; 1.  
 DR PROSITE: PS00985; SPERMADHESIN\_1; 1.  
 DR PROSITE: PS00986; SPERMADHESIN\_2; 1.  
 DR PROSITE: PS01180; CUB; 1.  
 DR Growth factor; Signal; 3D-structure.  
 KW SIGNAL  
 FT SIGNAL 1 20  
 FT CHAIN 21 134 ACIDIC SEMINAL FLUID PROTEIN.  
 FT DOMAIN 30 131 CUB.  
 FT DISULFID 30 51  
 FT DISULFID 74 95  
 FT CONFLICT 43 43 T -> H (IN REF. 2).  
 SQ SEQUENCE 134 AA; 15036 MW; 339BCFF8637D64C0 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 134;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 STYVDSMDY 10  
 DB 16 ATYVDSMDW 23  
 RESULT 7  
 ID GLB\_CERRH STANDARD; PRT; 151 AA.  
 AC P02215;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE Globin (Myoglobin).  
 OS Cerithiidae rhizopneurum (Water snail) (Horn shell).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;  
 OC Neotaeniolossea; Cerithioidae; Potamididae; Cerithiidae.  
 OK NCBI\_TaxID=6472;  
 RN (1)  
 RP SEQUENCE.  
 RC TISSUE=Radular muscle;  
 RX MEDLINE=83204919; PubMed=6849938;  
 RA Takagi T., Tobita M., Shikama K.;  
 RT "Amino acid sequence of dimeric myoglobin from Cerithiidae  
 rhizopneurum.";  
 RL Biochim. Biophys. Acta 745:32-36(1983).  
 CC -1- SUBUNIT: HOMODIMER.  
 DR PIR: A02537; GGGACR.  
 DR HSSP: P14821; 1SCR.  
 DR InterPro: IPR002336; Erythcrutin.  
 DR InterPro: IPR000971; Globin.  
 DR Pfam: PF00042; globin; 1.  
 DR PRINTS: PR00611; ERYTHCRUTIN.  
 DR PROSITE: PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle; Acetylation.  
 FT MOD\_RES 1 1  
 FT METAL 66 66 ACETYLATION.  
 FT METAL 98 98 IRON (HEME DISTAL LIGAND)  
 FT METAL 98 98 IRON (HEME PROXIMAL LIGAND)  
 FT METAL 98 98 (BY SIMILARITY).  
 SQ SEQUENCE 151 AA; 16210 MW; 3493BAE8F4A4BD90 CRC64;  
 Query Match 63.5%; Score 33; DB 1; Length 151;  
 Best Local Similarity 75.0%; Pred. No. 13;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 GSTVDSMD 9  
 DB 75 GSWDSMD 82  
 RESULT 8  
 ID MERS\_ARATH STANDARD; PRT; 247 AA.  
 AC P24806;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 DE MERI-5 protein.  
 DE MERI-5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OK NCBI\_TaxID=3702;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93005704; PubMed=1840916;

```

RA Medford J.I., Elmer J.S., Klee H.J.;
RT "Molecular cloning and characterization of genes expressed in shoot
RT apical meristems.";
RL Plant Cell 3:359-370(1991).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN SHOOT APICAL MERISTEMS, ALSO
CC FOUND IN SEEDLINGS AND MERISTEMS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63166; AAA32828.1; -.
DR PIR: J01022; J01022.
DR HSSP: P23904; IAKK.
DR InterPro: IPR000757; Glyco_hydro_16.
DR Pfam: PF00722; Glyco_hydro_16; 1.
SQ SEQUENCE 247 AA; 28295 MW; 831E8441564B45E8 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 247;
Best Local Similarity 60.0%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
DB 92 EGSTWDEIDF 101

RESULT 9
DHLE_BACCE STANDARD: PRT; 366 AA.
ID DHLE_BACCE
AC Q93194;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leucine dehydrogenase (EC 1.4.1.9) (LeudH).
GN LDH.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 626;
RA Stoyan T.; Recktenwald A.; Kula M.R.;
RT "Cloning, sequencing and overexpression of the leucine dehydrogenase
RT gene from Bacillus cereus.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FUNCTIONS CATABOLICALLY IN THE BACTERIAL METABOLISM OF
CC BRANCHED-CHAIN L-AMINO ACIDS, AND PLAYS AN IMPORTANT ROLE IN SPORE
CC GERMINATION IN COOPERATION WITH ALANINE DEHYDROGENASE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-leucine + H(2)O + NAD(+) = 4-methyl-2-
CC oxopentanoate + NH(3) + NADH.
CC -1- SUBUNIT: HOMOHExAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U51099; AAA96314.1; -.
DR HSSP: P80319; IGTM.
DR InterPro: IPR001625; GLFV_dehydrog.
DR Pfam: PF00208; GLFV_dehydrog; 1.

```

```

DR Pfam: PF02812; GLFV_dehydrog_N; 1.
DR PRINTS: PR00082; GLPFDHGNASE.
DR PROSITE: PS00074; GLFV_DEHYDROGENASE; 1.
RW Oxidoreductase; NAD: Branched-chain amino acid metabolism.
FT ACT_SITE 82 82 BY SIMILARITY.
FT NP_BIND 182 188 NAD (POTENTIAL).
SQ SEQUENCE 366 AA; 39867 MW; DA84E58062E772AC CRC64;

Query Match 63.5%; Score 33; DB 1; Length 366;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMD 9
DB 119 GTTVDDMD 126

RESULT 10
ECB2_HALEL STANDARD: PRT; 421 AA.
ID ECB2_HALEL
AC O52250;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Diaminobutyrate--pyruvate aminotransferase (EC 2.6.1.46) (L-
DE diaminobutyric acid transaminase) (Diaminobutyrate transaminase).
GN ECTP.
OS Halomonas elongata.
OC Bacteria; Proteobacteria; gamma subdivision; Halomonadaceae;
OC Halomonas.
OX NCBI_TaxID=2746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 25817;
RX MEDLINE=98231640; PubMed=9570121;
RA Goller K., Ofer A., Galinski E.A.;
RT "Construction and characterization of an NaCl-sensitive mutant of
RT Halomonas elongata impaired in ectoine biosynthesis.";
RL FEMS Microbiol. Lett. 161:293-300(1998).
CC -1- CATALYTIC ACTIVITY: L-2,4-diaminobutanoate + pyruvate = L-
CC aspartate 4-semialdehyde + L-alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF ECTOINE (1,4,5,6-
CC TETRAHYDRO-2-METHYL-4-PYRIMIDINE CARBOXYLIC ACID).
CC -1- SUBUNIT: HOMOHExAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF031489; AAC15882.1; -.
DR InterPro: IPR000954; Aminotran_3.
DR Pfam: PF00202; aminotran_3; 2.
DR PROSITE: PS00600; AA_TRANSFER_CLASS_3; FALSE NEG.
RW Transferase; Aminotransferase; Pyridoxal phosphate.
FT BINDING 267 267 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 421 AA; 46166 MW; AAA2E21596E1E16C CRC64;

Query Match 63.5%; Score 33; DB 1; Length 421;
Best Local Similarity 55.6%; Pred. No. 41;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10
DB 176 GSSTPDLDY 184

```

```

RESULT 11
YRVA_CAEEL STANDARD: PRT: 518 AA.
ID YRVA_CAEEL
AC 027519;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative cytochrome P450 CYP13A7 (EC 1.14.-.-).
GN CYP13A7 OR T10B9.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gardner A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 248717; CAA8609.1; -
CC WormPep: T10B9.10; CE01655.
CC InterPro: IPR001128; Cyt_P450.
CC Pfam: PF00067; P450.1.
CC PRINTS: PR00385; P450.1.
CC PROSITE: PS00086; CYTOCHROME_P450.1.
CC KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
CC FT BINDING 464 464 HEME (BY SIMILARITY).
CC SEQUENCE 518 AA; 58999 MW; CC04283EF87B9EAT CRC64;

Query Match 63.5%; Score 33; DB 1; Length 518;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10
DB 165 EGKTLDMLEY 174

RESULT 12
E13B_BACCI STANDARD: PRT: 682 AA.
ID E13B_BACCI
AC P23903;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glucan endo-1,3-beta-glucosidase A1 precursor (EC 3.2.1.39) ((1->3)-
DE beta-glucan endohydrolase) ((1->3)-beta-glucanase A1).
GN GlcA.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-52.
RC STRAIN-WL-12;
RC MEDLINE=90185240; PubMed=2311931;
RA Yahata N., Watanabe T., Nakamura Y., Yamamoto Y., Kamimiyu S.,
RA Tanaka H.;

```

```

RT "Structure of the gene encoding beta-1,3-glucanase A1 of Bacillus
RT circulans WL-12."
RL Gene 86:113-117(1990).
CC -1- FUNCTION: LYSIS OF CELLULAR WALLS CONTAINING BETA-1,3-GLUCANS.
CC IMPACTED IN THE DEFENSE AGAINST FUNGAL PATHOGENS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-glucosidic linkages
CC in 1,3-beta-D-glucans.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34503; AAA22474.1; -
CC PIR: J00420; J00420.
CC DR HSSP: P23904; IAOJ.
CC DR InterPro: IPR000757; Glyco_hydro.16.
CC DR Pfam: PF00722; Glyco_hydro.16; 1.
CC DR PROSITE: PS01034; GLYCOSYL_HYDROL_F16; 1.
CC KW Cell wall; Hydrolase; Glycosidase; Signal.
CC FT SIGNAL 1 38
CC FT CHAIN 1 38
CC FT ACT_SITE 39 682 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE A1.
CC FT ACT_SITE 552 552 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT_SITE 557 557 PROTON DONOR (BY SIMILARITY).
CC SEQUENCE 682 AA; 75465 MW; 8CAF407E34D4ADD5 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 682;
Best Local Similarity 66.7%; Pred. No. 71;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10
DB 40 GTTVISMEX 48

RESULT 13
PDC2_CANAL STANDARD: PRT: 836 AA.
ID PDC2_CANAL
AC O60035;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PDC2 protein.
GN PDC2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaiser B., Munder T., Kuenkel W., Eck R.;
RT "Molecular cloning and characterization of PDC2 from Candida
RT albicans."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR THE SYNTHESIS OF PYRUVATE DECARBOXYLASE
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y17007; CAA76580.1; -
CC FT DOMAIN 684 698 POLY-GLN.
CC FT DOMAIN 747 756 POLY-PRO.

```



SQ SEQUENCE 836 AA; 94360 MW; EA030FC8D5C694C7 CRC64;  
 Query Match 63.5%; Score 33; DB 1; Length 836;  
 Best Local Similarity 62.5%; Pred. No. 89;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 STVDSMDY 10  
 11:111  
 Db 817 STIDLDY 824

RESULT 14  
 ID CA16\_MOUSE STANDARD: PRT: 1025 AA.  
 AC 004857;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(VI) chain precursor.  
 GN COL6A1.  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93316904; PubMed=8326912;  
 RA Bonaldo P., Piccolo S., Marvulli D., Volpin D., Bressan G.M.;  
 RT "Murine alpha 1(VI) collagen chain. Complete amino acid sequence and  
 RT identification of the gene promoter region."  
 RL Matrix 13:223-233(1993).  
 RP [2]  
 RP SEQUENCE OF 442-1025 FROM N.A.  
 RX MEDLINE=93256888; PubMed=8489506;  
 RA Zhang R.Z., Pan T.C., Timpl R., Chu M.L.;  
 RT "Cloning and sequence analysis of cDNAs encoding the alpha 1, alpha 2  
 RT and alpha 3 chains of mouse collagen VI."  
 RL Biochem. J. 291:787-792(1993).  
 CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),  
 CC ALPHA 2(VI), AND ALPHA 3(VI).  
 CC -1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 3 VWFA DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X66405; CAA47032.1;  
 DR EMBL: X66406; CAA47033.1;  
 DR EMBL: Z18271; CAA79152.1;  
 DR PIR: S34839; S34839.  
 DR PIR: S31403; S31403.  
 DR MGI: 88459; Col6a1.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002035; VWFA.  
 DR Pfam: PF01391; Collagen. 5.  
 DR Pfam: PF00092; vwa; 3.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00327; VWA; 3.  
 DR PROSITE: PS50234; VWFA; 3.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 FT SIGNAL 1 19  
 FT CHAIN 20 1025 COLLAGEN ALPHA 1(VI) CHAIN.  
 FT DOMAIN 20 255 N-TERMINAL GLOBULAR DOMAIN.  
 FT DOMAIN 256 591 TRIPLE-HELICAL REGION.

FT DOMAIN 592 1025 C-TERMINAL GLOBULAR DOMAIN.  
 FT DOMAIN 36 234 VWFA 1.  
 FT DOMAIN 614 802 VWFA 1.  
 FT DOMAIN 826 1018 VWFA 3.  
 FT SITE 261 263 CELL ATTACHMENT SITE.  
 FT SITE 441 443 CELL ATTACHMENT SITE.  
 FT SITE 477 479 CELL ATTACHMENT SITE.  
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 893 893 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 674 675 DM -> TL (IN REF. 2).  
 FT CONFLICT 709 709 T -> A (IN REF. 2).  
 FT CONFLICT 943 943 MISSING (IN REF. 2).  
 FT CONFLICT 960 960 O -> R (IN REF. 2).  
 SQ SEQUENCE 1025 AA; 108489 MW; 2A05DFED8771BBF7 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 1025;  
 Best Local Similarity 75.0%; Pred. No. 1;le+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 3 STVDSMDY 10  
 1:11111  
 Db 899 SSVDSMDF 906

RESULT 15  
 ID ACYO\_MOUSE STANDARD: PRT: 98 AA.  
 AC P56376;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acylphosphatase, organ-common type isozyme (EC 3.6.1.7)  
 GN ACYP1 OR ACYPE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
 RA Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,  
 RA Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,  
 RA Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ITS PHYSIOLOGICAL ROLE IS NOT YET CLEAR.  
 CC -1- CATALYTIC ACTIVITY: An acyl phosphate + H(2)O -> a fatty acid anion  
 CC + phosphate.  
 CC -1- SIMILARITY: BELONGS TO THE ACYLPHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AA142466; ?; NOT\_ANNOTATED\_CDS.  
 DR EMBL: W83797; ?; NOT\_ANNOTATED\_CDS.  
 DR HSP: P41500; 2ACY.  
 DR InterPro: IPR001792; Acylphosphatase.  
 DR Pfam: PF00708; Acylphosphatase. 1.  
 DR PRINTS: PR00112; ACYLPHPTASE.  
 DR Prodom: PD001884; Acylphosphatase; 1.  
 DR PROSITE: PS00150; ACYLPHOSPHATASE\_1; 1.  
 DR PROSITE: PS00151; ACYLPHOSPHATASE\_2; 1.  
 KW Hydrolyase; Acetylase; Multigene family.  
 FT INIT\_MET 0 0 BY SIMILARITY.

FT	MOD_RES	1	1	ACETYLATION (BY SIMILARITY).
FT	ACT_SITE	23	23	POTENTIAL.
FT	ACT_SITE	41	41	POTENTIAL.
SQ	SEQUENCE	98 AA;	11110 MW;	A73893A2B6AA8911 CRC64;

Query Match 61.5%; Score 32; DB 1; Length 98;  
Best Local Similarity 60.0%; Pred. No. 13;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy	1	EGSTVDSMDY	10
	1	1	1
Db	2	EGDTLVSVDP	11

Search completed: August 20, 2002, 13:17:52  
Job time: 261 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:25 ; Search time 73.7 Seconds  
(without alignments)  
23.473 Million cell updates/sec

Title: US-09-824-286-3\_COPY\_95\_104  
Perfect score: 52  
Sequence: 1 EGSTVDSMDY 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39	75.0	5157	3	Q01135	Q01135 metarhizium
2	38	73.1	262	16	097019	097019 streptococcus
3	38	73.1	345	5	Q9VXN9	Q9VXN9 drosophila
4	38	73.1	372	10	Q94H16	Q94H16 Oryza sativa
5	37	71.2	329	16	Q98108	Q98108 rhizobium
6	37	71.2	1956	4	Q9Y5Y9	Q9Y5Y9 homo sapien
7	36	69.2	2005	5	Q9VFP5	Q9VFP5 drosophila
8	35	67.3	260	10	Q9STP2	Q9STP2 arabidopsis
9	35	67.3	280	5	Q18653	Q18653 caenorhabdi
10	35	67.3	484	5	Q9TZK6	Q9TZK6 caenorhabdi
11	35	67.3	535	2	Q68212	Q68212 actinomyces
12	35	67.3	544	4	Q9H087	Q9H087 homo sapien
13	35	67.3	621	4	Q94952	Q94952 homo sapien
14	35	67.3	1254	5	Q9VFS1	Q9VFS1 drosophila
15	34	65.4	117	2	Q9LAC2	Q9LAC2 uncultured
16	34	65.4	117	2	Q9LAC1	Q9LAC1 uncultured

17	34	65.4	117	2	09LAC0	09LAC0 uncultured
18	34	65.4	117	2	09L4B8	09L4B8 uncultured
19	34	65.4	117	2	09L4B6	09L4B6 uncultured
20	34	65.4	117	2	09L4B5	09L4B5 uncultured
21	34	65.4	117	2	09L4B4	09L4B4 uncultured
22	34	65.4	117	2	09L4B3	09L4B3 uncultured
23	34	65.4	117	2	09L4B1	09L4B1 uncultured
24	34	65.4	117	2	09L4A9	09L4A9 uncultured
25	34	65.4	117	2	09L4A8	09L4A8 uncultured
26	34	65.4	117	2	09L4A7	09L4A7 uncultured
27	34	65.4	117	2	09L4A6	09L4A6 uncultured
28	34	65.4	117	2	09L4A2	09L4A2 uncultured
29	34	65.4	117	2	09L4A1	09L4A1 uncultured
30	34	65.4	117	2	09L4A0	09L4A0 uncultured
31	34	65.4	117	2	09L4A9	09L4A9 uncultured
32	34	65.4	117	2	09L4A8	09L4A8 uncultured
33	34	65.4	233	2	093T99	093T99 azospirillum
34	34	65.4	233	2	093T97	093T97 azospirillum
35	34	65.4	257	16	09K0A6	09K0A6 vibrio chol
36	34	65.4	274	17	09YDH9	09YDH9 aeropyrum p
37	34	65.4	281	2	09ADV7	09ADV7 ehrlichia c
38	34	65.4	329	5	09TY08	09TY08 caenorhabdi
39	34	65.4	387	4	09HE7	09HE7 homo sapien
40	34	65.4	406	4	09H678	09H678 homo sapien
41	34	65.4	521	4	096SP3	096SP3 homo sapien
42	34	65.4	544	10	09FIF7	09FIF7 arabadopsis
43	34	65.4	831	4	09NVY8	09NVY8 homo sapien
44	34	65.4	896	4	09BVJ1	09BVJ1 homo sapien
45	33.5	64.4	303	16	097RB6	097RB6 streptococc

## ALIGNMENTS

RESULT 1  
001135 PRELIMINARY: PRT; 5157 AA.  
AC 001135;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PEPTIDE SYNTHETASE.  
GN PESA.  
OS Metarhizium anisopliae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.  
OX NCBI\_TaxID=5350;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MEL;  
RX MEDLINE=97082966; PubMed=8964498;  
RA Bailey A.M., Kershaw M.J., Hunt B.A., Patterson I.C., Charnley A.K.,  
RA Reynolds S.E., Clarkson J.M.;  
RT "Cloning and sequence analysis of an intron containing domain of a  
RT peptide synthetase from the entomopathogenic fungus Metarhizium  
RT anisopliae.";  
RL Gene 173:195-197(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MEL;  
RA Bailey A.M., Reynolds S.E., Charnley A.K., Clarkson J.M.;  
RT "Evidence for multiple peptide synthetases from Metarhizium  
RT anisopliae.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: X89442; CA61605.1; -;  
DR HSSP: P14687; IAMU.  
DR InterPro: IPR000873; AMP-bind.  
DR InterPro: IPR001242; DUF4.  
DR InterPro: IPR003880; Phosphonat\_attach.  
DR Pfam: PF00501; AMP-binding; 4.  
DR Pfam: PF00668; Condensation; 7.  
DR Pfam: PF00550; PP-binding; 4.  
DR PRINTS: PR00154; AMPBINDING.

DR PROSITE: PS50075; ACP\_DOMAIN; 4.  
DR PROSITE: PS00455; AMP\_BINDING; 4.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
KM Phosphopantetheine.  
SQ SEQUENCE 5157 AA; 573954 MW; 1038242BA3143868 CRC64;

Query Match 75.0%; Score 39; DB 3; Length 5157;  
Best Local Similarity 77.8%; Pred. No. 2.4e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9  
DB 2003 EGSTVDSMD 2011

RESULT 2  
ID 097019 PRELIMINARY; PRT: 262 AA.  
AC 097019;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PROIPOPOTEIN DIACYLGLYCERL TRANSFERASE.  
GN SP1412.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_Taxid=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TIG84;  
RA MEDLINE-21357209; PubMed-11463916;  
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayar L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapfel E., Kouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae."  
RT Science 293:498-506(2001).  
RL EMBL: AF007438; AAK75510.1; -.  
DR TIGR: SP1412; -.  
DR TIGR: SP1412; -.  
DR InterPro: IPR001640; LGT.  
DR Pfam: PF01790; LGT; 1.  
DR PROSITE: PS01311; LGT; 1.  
KW Transferase; Lipoprotein; Complete proteome.  
KW SEQUENCE 262 AA; 30292 MW; B3A50C5FA8BDC19 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 262;  
Best Local Similarity 66.7%; Pred. No. 15;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSTVDSMDY 10  
DB 142 GATVDNDLY 150

RESULT 3  
ID 09VXN9 PRELIMINARY; PRT: 345 AA.  
AC 09VXN9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CG895 PROTEIN (PEPTIDOGLYCAN-RECOGNITION PROTEIN-LE) (GH01554P).  
GN PGRP-LE OR CG895.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RA MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaniatis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers V.-H.C., Blazek R.G., Champs M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhanderi D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Zhang G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RT Science 287:2185-2195(2000).  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20558582; PubMed-11106397;  
RA Werner T., Liu G., Kang D., Ekengren S., Steiner H., Hultmark D.,  
RT "A family of peptidoglycan recognition proteins in the fruit fly  
RT Drosophila melanogaster."  
RT Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).  
RL [3]  
RN SEQUENCE FROM N.A.  
RP STRAIN-Y, CN BW SP;  
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champs M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunco J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
RL EMBL: AE003500; AAF48519.1; -.  
DR EMBL: AF313391; AAG32064.1; -.  
DR EMBL: AT058258; AAL15487.1; -.  
DR Flybase: FBgn0030695; PGRP-LE.  
SQ SEQUENCE 345 AA; 39426 MW; 66CC484B54705AD7 CRC64;

Query Match 73.1%; Score 38; DB 5; Length 345;  
Best Local Similarity 70.0%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10

Db 1 1 1 1 1 1  
31 ESSYVDSLDY 40

RESULT 4  
ID Q94HI6 PRELIMINARY; PRT: 372 AA.  
AC Q94HI6;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHEtical PROTEIN.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wang R.A., Fisch D., Presting G., Wood T., Yu Y., Soderlund C.,  
RA Kim H.-R., Rambo T., Henry D., Simmons J.;  
RT "Rice Genomic Sequence."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC079632; AAK63901.1; -  
SQ SEQUENCE 372 AA; 42154 MW; D9F3C3822D5A2AB CRC64;

Query Match 73.1%; Score 38; DB 10; Length 372;  
Best Local Similarity 66.7%; Pred. No. 22;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9  
Db 29 EGATIDNMD 37

RESULT 5  
ID Q981U8 PRELIMINARY; PRT: 329 AA.  
AC Q981U8;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE MLP228 PROTEIN.  
GN MLP228.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-MAFF303099;  
RC MEDLINE-21082930; PubMed-11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpoto S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti."  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003015; BAB54611.1; -  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 329 AA; 35675 MW; 86CEACBBI55345DC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 329;  
Best Local Similarity 60.0%; Pred. No. 30;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMDY 10  
1 1 1 1 1 1 1 1

Db 207 EGSTVETVDY 216

RESULT 6  
ID Q9Y5Y9 PRELIMINARY; PRT: 1956 AA.  
AC Q9Y5Y9;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TETRODOTOXIN-RESISTANT VOLTAGE-GATED SODIUM CHANNEL.  
GN SCN10A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-DORSAL ROOT GANGLION;  
RX MEDLINE-9905355; PubMed-9839820;  
RA Robert D.K., Koch B.D., Illicka M., Obernolte R.A., Naylor S.L.,  
RA Herman R.C., Eglen R.M., Hunter J.C., Sangameswaran L.;  
RT "A tetrodotoxin-resistant voltage-gated sodium channel from human  
dorsal root ganglia, hPNS/SCN10A."  
RL Pain 78:107-114(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-DORSAL ROOT GANGLION;  
RA Robert D.K., Koch B.D., Illicka M., Obernolte R.A., Naylor S.L.,  
RA Herman R.C., Eglen R.M., Hunter J.C., Sangameswaran L.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF117907; AAD30863.1; -  
DR InterPro; IPR000636; Catlon\_chan\_non\_lig.  
DR InterPro; IPR002111; Cat\_channel\_Tryp.  
DR InterPro; IPR001682; Channel\_pore\_Ca\_Na.  
DR InterPro; IPR000048; IQ.  
DR InterPro; IPR001696; Na\_channel.  
DR Pfam; PF00520; Ion\_Trans\_4.  
DR Pfam; PF00612; IQ\_1.  
DR PRINTS; PR00170; NACHANNEL.  
KW Ionic channel.  
SQ SEQUENCE 1956 AA; 220564 MW; F24B73E7C211DA5E CRC64;

Query Match 71.2%; Score 37; DB 4; Length 1956;  
Best Local Similarity 77.8%; Pred. No. 2.1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGSTVDSMD 9  
Db 1085 EGSTVDCID 1093

RESULT 7  
ID Q9VFN5 PRELIMINARY; PRT: 2005 AA.  
AC Q9VFN5;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PUTATIVE CADHERIN PRECURSOR (CG3389 PROTEIN).  
GN CG3389.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-BERKELEY;  
RC MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Suttton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson S.C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Splter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,  
 RA Zhang X.H., Myers E.W., Rubin G.M., Venter J.C., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Zeng F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 Science 287:2185-2195(2000).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC EMBL: AE003705; AAF5082.1; -.  
 DR HSP: P15116; INCI.  
 DR FlyBase: FBgn0038247; CG3389.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 14.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 14.  
 DR PROSITE: PS00232; CADHERIN\_1; 1.  
 DR PROSITE: PS00268; CADHERIN\_2; 14.  
 KW Hypothetical protein; Cell adhesion; signal; Transmembrane; Repeat;  
 KW Glycoprotein; Calcium-binding  
 FT SIGNAL 1 55 POTENTIAL.  
 FT CHAIN 56 2005 POTATIVE CADHERIN.  
 FT DOMAIN 56 1849 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1850 1870 POTENTIAL.  
 FT DOMAIN 1871 2005 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 75 172 CADHERIN 1.  
 FT DOMAIN 181 281 CADHERIN 2.  
 FT DOMAIN 290 393 CADHERIN 3.  
 FT DOMAIN 402 510 CADHERIN 4.  
 FT DOMAIN 519 621 CADHERIN 5.  
 FT DOMAIN 630 746 CADHERIN 6.  
 FT DOMAIN 755 880 CADHERIN 7.  
 FT DOMAIN 888 1001 CADHERIN 8.  
 FT DOMAIN 1010 1105 CADHERIN 9.  
 FT DOMAIN 1114 1219 CADHERIN 10.  
 FT DOMAIN 1228 1328 CADHERIN 11.  
 FT DOMAIN 1339 1438 CADHERIN 12.  
 FT DOMAIN 1447 1629 CADHERIN 13.  
 FT DOMAIN 1638 1743 CADHERIN 14.  
 FT CARBOHYD 431 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 516 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 516 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 867 867 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1111 1111 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1334 1334 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1342 1342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1446 1446 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1516 1516 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1529 1529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1718 1718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1764 1764 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 2005 AA: 224474 MW: 6482A1D19C3A3C47 CRC64;

Query Match Score 36; DB 5; Length 2005;  
 Best Local Similarity 66.7%; Pred No. 3.5e-02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 2 GSTVDSMDY 10  
 Db 11 GNGIDSMXD 19

RESULT 8  
 ID 09STP2 PRELIMINARY; PRT; 260 AA.  
 AC 09STP2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 13, Last annotation update)  
 DE HYPOTHETICAL 29.2 KDA PROTEIN.  
 GN T27E11.80 OR AT4G27840.  
 DN Arabidopsis thaliana (Mouse-ear cress).  
 OS Arabidopsis: Viridiplantae: Streptophyta: Tracheophyta;  
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;  
 OC Eurosids II: Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,  
 RA Barrell B.G., Bancroft I., Mewes H.W., Meyer K.F.X.,  
 RA Schellier C.,  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL078579; CAB43972.1; -.  
 DR EMBL: AL161571; CAB41433.1; -.  
 KW Hypothetical protein  
 SQ SEQUENCE 260 AA: 29203 MW: 6482A1D19C3A3C47 CRC64;

Query Match Score 35; DB 10; Length 260;  
 Best Local Similarity 66.7%; Pred No. 59;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 EGSTVDSMD 9  
 Db 98 DSTADSLD 106

```

RESULT 9
ID 018663 PRELIMINARY; PRT: 280 AA.
AC 018663;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHETICAL 31.5 KDA PROTEIN.
GN C46H3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; Pubmed=9851916;
RA None;
RT "genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A.;
RT "The sequence of C. elegans cosmid C46H3.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL EMBL: U41271; AAA82445.1; -.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank; 6.
DR SMART: SM00248; ANK; 5.
DR PROSITE: PS50088; ANK_REPEAT; 3.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Hypothetical protein; Repeat.
SQ SEQUENCE 280 AA; 31530 MW; 254EBA4EDE033866 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 280;
Best Local Similarity 66.7%; Pred. No. 64;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10
Db 130 GANVDSVDY 138
1:|||||
O9TZK6 PRELIMINARY; PRT: 484 AA.
AC 09TZK6;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE C36C9.5 PROTEIN.
GN C36C9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

```

```

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Swadlow N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton B., Martin J., O'Brien D.;
RT "The sequence of C. elegans cosmid C36C9.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF098986; AAC67427.1; -.
SQ SEQUENCE 484 AA; 55327 MW; CCC1ECBBE82C441 CRC64;

```

```

Query Match 67.3%; Score 35; DB 5; Length 484;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
Db 235 EGSSEETMDY 244
|||||
O68212 PRELIMINARY; PRT: 535 AA.
AC 068212;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE FIMBRIAL STRUCTURAL SUBUNIT.
GN FIMA.
OS Actinomyces naeslundii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_Taxid=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T14V;
RX MEDLINE=98187920; Pubmed=9529071;
RA Yeung M.K., Donkersloot J.A., Cisar J.O., Ragsdale P.A.;
RT "Identification of a gene involved in assembly of Actinomyces
RT naeslundii T14V type 2 fimbriae.";
RL Infect. Immun. 66:1482-1491(1998).
DR EMBL: AF019629; AAC13545.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 535 AA; 56038 MW; B115867F500C1356 CRC64;

```

```

Query Match 67.3%; Score 35; DB 2; Length 535;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 EGSTVDSMDY 10
Db 263 EGSTLDPTDY 272
|||:|:|
O9H087 PRELIMINARY; PRT: 544 AA.
ID 09H087

```

AC Q9H087;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHEICAL. 63.3 KDA PROTEIN.  
GN DKFZP346G058.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
RA Ansgore W., Boecker M., Bloecher H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
RA Mewes H.W., Oltmair B., Obermayer B., Tampe J., Heubner D.,  
RA Wandt R., Korn B., Klein M., Poustka A.;  
RT Towards a Catalog of Human Genes and Proteins: Sequencing and  
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.;  
RL Genome Res. 11:422-435(2001).  
DR EMBL; AL136899; CAB66833.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 544 AA; 63306 MW; D95BE3CEA3B7D0ED CRC64;

Query Match 67.3%; Score 35; DB 4; Length 544;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EGSTVDSMDY 10  
11:1:1 11  
Db 253 EGATLDIFDY 262

RESULT 13  
Q94952 PRELIMINARY; PRT; 621 AA.  
AC Q94952;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE KIA0875 PROTEIN (FRAGMENT).  
GN KIA0875 OR FBX21.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRIN;  
RA MEDLINE=99156230; PubMed=10048485;  
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,  
RA Miyajima N., Tanaka K., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:355-364(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=20003060; PubMed=10531035;  
RA Cenciarelli C., Chisaur D.S., Guardavaccaro D., Parks W., Vidal M.,  
RA Pagano M.;  
RT "Identification of a family of human F-box proteins.";  
RL Curr. Biol. 9:1177-1179(1999).  
DR EMBL; AB020682; BAA74898.1; -  
DR EMBL; AF14601; AAF04522.1; -  
DR InterPro; IPR001810; F-box.  
DR Pfam; PF00646; F-box; 1.  
FT NON\_TER 1  
SQ SEQUENCE 621 AA; 71479 MW; D302DD46F3ABB5A3 CRC64;

Query Match 67.3%; Score 35; DB 4; Length 621;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EGSTVDSMDY 10  
11:1:1 11  
Db 337 EGATLDIFDY 346

RESULT 14  
Q9VF51 PRELIMINARY; PRT; 1254 AA.  
AC Q9VF51;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE CG6045 PROTEIN.  
GN CG6045.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burlis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dudon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gilbert W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Schelder F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradlin A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -I- COPACTOR: BINDS A 2PE-2S CLUSTER (BY SIMILARITY).  
DR EMBL; AE003709; AAF55209.1; -  
DR FlyBase; FBgn0038349; CG6045.  
DR InterPro; IPR002888; 2Pe-2S\_BD.  
DR InterPro; IPR000564; 2Pe2S\_ferredoxin.  
DR InterPro; IPR000674; Aldxan\_dh.C.  
DR InterPro; IPR002346; dehydrog\_molylb.



DR InterPro: IPR001041; Ferredoxin.  
 DR Pfam: PF01315; Ald\_Xan\_dh\_C; 2.  
 DR Pfam: PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam: PF00941; FAD\_binding\_5; 1.  
 DR Pfam: PF00111; fer2; 1.  
 DR Pfam: PF01789; fer2\_2; 1.  
 DR ProDom: PD186071; 2Fe-2S\_BD; 1.  
 DR PROSITE: PS00197; 2FE2S\_FERREDOXIN; 1.  
 KW Iron-sulfur.  
 SO SEQUENCE 1254 AA; 137886 MW; 4CC7E24F15D80851 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 1254;  
 Best Local Similarity 75.0%; Pred. No. 3.3e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSTVDSMD 9  
 |||:|:|  
 DB 567 GSTIDSID 574

RESULT 15  
 O9LAC2  
 ID O9LAC2 PRELIMINARY; PRT; 117 AA.  
 AC O9LAC2;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE NITRITE REDUCTASE (FRAGMENT).  
 GN NIRS.  
 OS uncultured bacterium PA4.  
 OC Bacteria: environmental samples.  
 OX NCBI\_TaxID=112731;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20250661; PubMed=10788387;  
 RA Braker G., Zhou J., Wu L., Devol A.H., Tiedje J.M.;  
 RT "Nitrite reductase genes (nirK and nirS) as functional markers to  
 RT investigate diversity of denitrifying bacteria in pacific northwest  
 RT marine sediment communities";  
 RL Appl. Environ. Microbiol. 66:2096-2104(2000).  
 DR EMBL: AJ248402; CAB6767.1; -.  
 DR HSSP: P24474; INIR.  
 FT NON\_TER 1  
 FT NON\_TER 117  
 SQ SEQUENCE 117 AA; 12923 MW; C8BD79385BC2776 CRC64;

Query Match 65.4%; Score 34; DB 2; Length 117;  
 Best Local Similarity 66.7%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSTVDSMDY 10  
 | | | | | | | | | |  
 DB 11 GMTVDTQDY 19

Search completed: August 20, 2002, 13:17:26  
 Job time: 240 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:24:06 ; Search time 49.92 Seconds

(without alignments)  
24.475 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_24\_34

Sequence: 1 KASQDVTAVV 11

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	106	19	AAW31647
2	46	92.0	250	17	AAW02278
3	46	92.0	250	18	AAW29261
4	46	92.0	250	18	AAW22400
5	46	92.0	250	19	AAW53168
6	46	92.0	250	19	AAW47012
7	46	92.0	250	20	AAW0422
8	46	92.0	667	14	AAW39573
9	45	90.0	11	16	AAW4946
10	45	90.0	11	19	AAW62189
11	45	90.0	11	19	AAW57577

12	45	90.0	11	20	AAW02547
13	45	90.0	11	20	AAW69629
14	45	90.0	11	21	AAW77515
15	45	90.0	11	22	AAW67113
16	45	90.0	11	22	AAW63395
17	45	90.0	11	22	AAW64777
18	45	90.0	11	22	AAW76891
19	45	90.0	11	22	AAW76910
20	45	90.0	11	22	AAW76929
21	45	90.0	107	16	AAW74958
22	45	90.0	109	14	AAW30766
23	45	90.0	126	19	AAW77293
24	45	90.0	126	19	AAW65772
25	45	90.0	126	19	AAW62208
26	45	90.0	126	19	AAW62195
27	45	90.0	126	20	AAW32769
28	45	90.0	126	20	AAW02553
29	45	90.0	126	20	AAW02554
30	45	90.0	126	20	AAW05485
31	45	90.0	130	16	AAW74968
32	45	90.0	131	19	AAW62187
33	45	90.0	131	20	AAW02545
34	45	90.0	11	16	AAW85774
35	44	88.0	107	16	AAW74959
36	44	88.0	120	16	AAW74969
37	43	86.0	108	20	AAW21817
38	42	84.0	11	19	AAW47101
39	42	84.0	109	14	AAW30762
40	42	84.0	109	15	AAW47039
41	42	84.0	237	13	AAW24047
42	42	84.0	252	21	AAW28158
43	42	84.0	513	22	AAW82924
44	42	84.0	698	20	AAW83493
45	40	80.0	11	17	AAW97313

#### ALIGNMENTS

RESULT 1	
ID	AAW31647 standard; Protein; 106 AA.
XX	
AC	AAW31647;
XX	
DT	21-MAY-1998 (first entry)
XX	
DE	Monoclonal antibody CP.B8 light chain variable region.
XX	
XX	
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; CP.B8; immunological disease;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW	grit versus host disease; psoriasis; immunosuppressive; therapy;
XX	complementarily determining region; CDR.
XX	
OS	Mus musculus.
XX	
PH	Key
FT	Region
FT	/note= "CDR1"
FT	/note= "CDR2"
FT	/note= "CDR3"
FT	Region
XX	
PN	MO9743416-AL.
XX	
PD	20-NOV-1997.
XX	
PF	09-MAY-1997; 97WO-US07870.
XX	

Artificial CDR(1)  
Mouse humanised an  
Antibody H chain V  
Amino acid sequenc  
Anti-PRRP Ab VL C  
Human PRRP mouse  
Human PRRP mouse  
Human PRRP mouse  
Human PRRP mouse  
Immunoglobulin lig  
Light chain variab  
Anti-human HMI.24  
Humanised anti-HMI  
Human anti-HMI.24  
Anti-HMI.24 antibo  
Humanised L chain  
Humanised L chain  
H chain V region o  
Anti-idiotypic anti  
Mouse anti-HMI.24  
L chain V region o  
L-CDR-1 of anti-id  
Immunoglobulin lig  
Anti-idiotypic anti  
Anti-STX1 light ch  
NMDA receptor bind  
Light chain variab  
Sequence of the 11  
Light chain of 4D5  
Human anti-C-ErbB2  
Anti-HER2 HuMab4D5  
4D5 Fab molecule e  
Humanised monocl

10-MAY-1996; 96US-0017466.  
(BIOJ ) BIOGEN INC.  
Benjamin CD, Burkiy LC, Hession C, Whitley A;  
WPI: 1998-008885/01.  
N-PSDB; AAT97440.  
Blocking agents of the gamma common chain of cytokine receptors -  
particularly monoclonal antibodies, used to induce T cell anergy for  
treatment of immunological diseases  
Claim 22; Page 81; 111pp; English.  
This polypeptide comprises the light chain variable region (VL) of  
monoclonal antibody (Mab) C9-B8, which is produced by a hybridoma  
cell deposited as ATCC 12107, and which is specific for the gamma  
constant (gc) chain (see AAM31646) of human cytokine receptors. The  
invention provides compositions and methods for inhibiting cytokine  
signalling using gc chain blocking agents for the treatment of  
immunological diseases such as myasthenia gravis, rheumatoid  
arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
allergy, asthma, parasitic infection, graft vs. host disease or  
psoriasis. Preferred gc blocking agents include Mab CP-B8, its Fab  
fragment and an antibody having a light chain variable region  
CDR selected from those of CP-B8 VL or a heavy chain variable  
region CDR selected from those of CP-B8 VH (see AAM31648).  
Sequence 106 AA;

Query Match	100.0%;	Score 50;	DB 19;	Length 106;
Best Local Similarity	100.0%;	Pred. No. 0.057;		
Matches 11;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 KASQDVTTPAVA 11			
	24 KASQDVTTPAVA 34			
Db				
RESULT 2				
AAW02278				
ID	AAW02278	standard; Protein: 250 AA.		
XX	AAW02278;			
AC				
XX				
D7	29-OCT-1996	(first entry)		
XX				
DE	741F8 anti-c-erbB-2	two single chain Fv construct.		
XX				
KW	741F8; anti-c-erbB-2; monoclonal antibody; single chain Fv; sfv;			
KW	construct; polypeptide linker; C-terminal amino acid sequence;			
KW	in vivo imaging; drug targeting experiment; homodimer;			
KW	increased; binding avidity; tissue retention time.			
XX				
CS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	Peptide	122..135		
FT	Peptide	/label= linker		
FT	Peptide	246..250		
FT	/note= "claimed C-terminal tail to facilitate			
FT	crosslinking of two sfv polypeptides"			
XX				
PN	US5534254-A.			
XX				
PD	09-JUL-1996.			
XX				
PF	06-FEB-1992;	92US-0831967.		
XX				
PR	07-OCT-1993;	93US-0133804.		

XX	06-FEB-1992:	92US-0831967.
XX	(CHIR ) CHIRON CORP.	
PA	(CREA-) CREATIVE BIOMOLECULES INC.	
XX		
XX	Houston LL, Huston JS, Oppermann H, Ring DB;	
PI		
PI	WPI: 1996-333194/33.	
DR	DR N-PSDB; AAT36878.	
PT	Compns. contg. antigen-targetting antibody fragment constructs	
XX	comprising dimer of single-chain Fv fragments	
XX	Claim 25; Columns 27-28; 30pp; English.	
XX		
CC	The variable heavy (VH) and variable light (VL) genes of the 7A1F8	
CC	anti-c-erbB-2 monoclonal antibody (MAb), were isolated from the	
CC	cDNA of the parental 7A1F8 hybridoma line. A two single chain Fv	
CC	(scFv) gene was constructed by connecting the VH and VL genes with a	
CC	DNA sequence encoding a polypeptide linker. A synthetic DNA duplex	
CC	encoding the C-terminal amino acid sequence, (gly)4-Cys was	
CC	inserted, and the resulting 7A1F8 anti-c-erbB-2 two scFv inserted	
CC	into an expression vector. The resulting gene, which encodes the	
CC	present sequence, was transformed into E. coli, and protein	
CC	expression induced by the addn. of IPTG to the culture medium.	
CC	A compsn. comprising a carrier and the 2 scFv protein prod. can be	
CC	used for in vivo imaging, and drug targetting experiments. The	
CC	2 scFv protein prod. is a homodimer, in which both fragments target	
CC	the same antigen, therefore giving greater binding avidity and	
CC	longer tissue retention times, compared to individual scFv protein	
CC	prod. fragments.	
XX		
XX		
XX	Sequence 250 AA:	
XX		

```

92.0%; Score 46; DB 17; Length 250;
Best Query Match 90.9%; Pred. No. 0.75;
Best Local Similarity 90.9%;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KASQDVTTAVA 11
|||||:||||
Db 159 kasqdvstava 169
.

RESULT 3
AAW29261
ID AAW29261 standard; Protein; 250 AA.
XX
AC AAW29261;
XX
DT 07-JAN-1998 (first entry)
XX
DE Anti-c-erbB-2 single chain antibody 741F8.
XX
KW Adenovirus; E1A; transactivator; transcription activator; stimulate;
KW expression vector; single-chain binding protein; VAI; enhance; PCR;
KW translocation; production; immortal; eukaryotic cell; scfv; primer;
KW single-chain antibody fragment; imaging; tumour; breast cancer;
KW ovarian cancer; c-erbB-2 antigen; digoxin intoxication.
XX
XX Synthetic.
OS
XX
PN US5658763-A.
XX
PD 19-AUG-1997.
XX
XX
PF 25-OCT-1993; 93US-0143498.
XX
PR 25-OCT-1993; 93US-0143498.
PR 05-JUN-1995; 95US-0463675.
XX
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX
PA
XX

```

PI Doral H, Oppermann H;  
 XX WPI: 1997-424235/39.  
 DR N-PSDB; AAT91837.

PT Producing single chain binding protein in immortalised eukaryotic  
 PT cells - which comprise protein coding sequences, a transcription  
 PT activator and translation promotion sequences, provides high  
 PT expression at low copy number

PS Example 2; Column 29-32; 24pp; English.

XX This sequence is a single chain anti-c-erbB-2 antibody (scFv)  
 CC derived from hybridoma 741F8. The protein product includes a C-terminal  
 CC tail Gly4-Cys in some constructs. Single-chain antibodies can be  
 CC produced using a novel method which comprises culturing an immortalised  
 CC eukaryotic cell having transfected DNA sequences (encoding the protein  
 CC of interest), integrated into its genome. In particular expression  
 CC effector vectors containing a non-native reporter DNA (encoding the scFv)  
 CC and viral sequences to promote transcription and translation (e.g. the  
 CC adenovirus E1A and VAI genes as shown in AAT91831 and AAT91834  
 CC respectively) are used. The scFv that is produced, when properly folded,  
 CC has a structure with mono- or bi-functional binding activity. The method  
 CC is especially used to produce single-chain antibody fragments (scFv),  
 CC e.g. for imaging tumours or delivering therapeutic agents to them,  
 CC particularly breast and ovarian cancers that express the c-erbB-2  
 CC antigen. Other scFv are used in model studies and for treating digoxin  
 CC intoxication.

SQ Sequence 250 AA;

Query Match 92.0%; Score 46; DB 18; Length 250;  
 Best Local Similarity 90.9%; Pred. No. 0.75;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTTAVA 11  
 |||||:||||  
 Db 159 kasqdvstava 169

RESULT 4  
 AAW22400 standard; Protein; 250 AA.  
 XX AAW22400:  
 AC 09-OCT-1997 (first entry)  
 DT 09-OCT-1997 (first entry)  
 DE Single chain antibody 741F8 protein sequence.  
 XX  
 KW Production: single-chain; binding protein; antibody; eukaryote; virus;  
 KW transcription activator; promoter; expression; adenovirus; E1A; PCR;  
 KW polymerase chain reaction; amplification; primer; herpes simplex virus;  
 KW thymidine kinase; vector; enhancer; translation; heterologous.  
 XX  
 OS Synthetic.  
 XX  
 PN US5631158-A.  
 PD 20-MAY-1997.  
 XX  
 PF 25-OCT-1993; 93US-0143498.  
 XX  
 PR 25-OCT-1993; 93US-0143498.  
 PR 05-JUN-1995; 95US-0461184.  
 XX  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PI Doral H, Oppermann H;  
 XX WPI: 1997-288577/26.  
 DR N-PSDB; AAT78879.

XX  
 PT Production of cell line for producing single-chain binding protein -  
 PT using construct containing DNA encoding viral transcription  
 PT activator protein

PS Example 2; Column 29-32; 24pp; English.

XX The invention relates to methods of increasing production of a  
 CC single-chain binding protein, especially a single chain antibody,  
 CC by generating eukaryotic cell lines containing DNA encoding either  
 CC a viral transcription activator protein that acts on and stimulates  
 CC a viral promoter controlling the expression of DNA encoding the  
 CC single-chain binding protein, such as the adenovirus Ad2 E1A protein  
 CC encoded by the sequence AAT78873, or an RNA sequence able to promote  
 CC translation of the RNA transcript from the heterologous gene, such as  
 CC the adenoviral VAI gene (AAT78876).  
 CC The sequence presented here is the amino acid sequence of the  
 CC anti-c-erbB2 single chain antibody 741. The coding sequence was  
 CC constructed by amplifying the variable heavy and light chain genes  
 CC which were then cloned into a pUC vector. For secretion from mammalian  
 CC cells, the sequence is preceded by the signal peptide sequence from  
 CC the monoclonal antibody 520C9 (AAT78880) or from Pac1 (AAT78881).

SQ Sequence 250 AA;

Query Match 92.0%; Score 46; DB 18; Length 250;  
 Best Local Similarity 90.9%; Pred. No. 0.75;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTTAVA 11  
 |||||:||||  
 Db 159 kasqdvstava 169

RESULT 5  
 AAW53168 standard; Protein; 250 AA.  
 XX AAW53168:  
 AC 16-JUL-1998 (first entry)  
 DT 16-JUL-1998 (first entry)  
 DE 741F8 anti-c-erbB-2 sfv' dimeric construct protein sequence.  
 XX  
 KW Antigen imaging; single chain Fv; sfv; linker; dimeric; cancer;  
 KW c-erbB-2; tumour; diagnosis.  
 XX  
 OS Synthetic.  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 122..135  
 FT Peptide /note="Linker peptide"  
 FT Peptide 246..250  
 FT /note="Gly4-Cys C-terminal tail"  
 XX  
 EN US5753204-A.  
 XX  
 PD 19-MAY-1998.  
 XX  
 PF 05-JUN-1995; 95US-0461838.  
 XX  
 PR 07-OCT-1993; 93US-0133804.  
 PR 06-FEB-1992; 92US-0831967.  
 PR 05-JUN-1995; 95US-0461838.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PI Houston LL, Huston JS, Oppermann H, Ring DB;  
 XX WPI: 1998-311318/27.



CC 99m-technetium) to antigen-expressing cells, particularly for treatment  
CC or diagnosis of tumours (especially of ovary or breast).

SO Sequence 250 AA;

Query Match 92.0%; Score 46; DB 20; Length 250;  
Best Local Similarity 90.9%; Pred. No. 0.75;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11  
| | | | | | | | | |  
DB 159 kasqdvstava 169

#### RESULT 8

AAK39573 8  
ID AAK39573 standard; Protein; 667 AA.

AC AAK39573;

DT 07-FEB-1994 (first entry)

DE Sequence of 741 SFV-PE40.

KW Single chain Fv polypeptide; VH-VL heterodimer; immunoglobulin; Ig;  
exotoxin.

OS Pseudomonas.

PN MO9316185-A.

PD 19-AUG-1993;

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

PT (CETU ) CETUS ONCOLOGY CORP.

PA (CREA-) CREATIVE BIOMOLECULES INC.

PI Houston LL, Huston JS, Oppermann H, Ring DB;

DR WPI; 1993-272889/34.

XX N-PSDB; AAQ46088.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour  
antigen - for imaging or treating breast or ovarian cancer etc.

PS Example; pages 71-74; 87pp; English.

CC C-erbB-2 refers to a protein antigen expressed on the surface of  
CC tumour cells, such as breast and ovarian tumour cells, which is an  
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
CC pt. of about 5.3 (see AAQ46083, AAK39568). A single chain Fv (SFV)  
CC is a covalently linked VH-VL heterodimer which is expressed from  
CC a gene fusion including VH- and VL- encoding genes connected by  
CC a peptide-encoding linker. Such linker sequences are set forth in  
CC AA residues 116-135 in AAK39569, which includes part of the 16 AA  
CC linker sequences in AAK39572. Using AAQ46084 for the 520C9 monoclonal  
CC antibody, a single chain polypeptide can be produced having a  
CC binding affinity for a c-erbB-2 related antigen. In another design,  
CC the pseudomonal exotoxin fragment analogous to ricin A chain, PE40,  
CC is fused to the carboxy terminus of the antic-c-erbB-2 741 SFV.  
CC the resulting 7418 SFV-PE40 is a single-chain Fv-toxin fusion  
CC protein, which was constructed with an 18 residue short FB leader  
CC which initially was left on the protein. 'X' in AAK39573 refers to  
CC posn. of stop codon in cDNA.

SO Sequence 667 AA;

Query Match 92.0%; Score 46; DB 14; Length 667;

Best Local Similarity 90.9%; Pred. No. 2;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11  
| | | | | | | | | |  
DB 159 kasqdvstava 169

#### RESULT 9

AAK74946 9  
ID AAK74946 standard; peptide; 11 AA.

AC AAK74946;

DT 19-JAN-1996 (first entry)

DE L-CDR-1 of anti-idiotype antibody against human anticancer antibody.

KW Antibody; cancer; CDR; heavy chain; light chain; immunoglobulin;  
complementarity determining region.

OS Mus sp.

PN JP07101999-A.

PD 18-APR-1995.

PF 06-OCT-1993; 93JP-0272950.

PR 06-OCT-1993; 93JP-0272950.

PT (HAGI/) HAGIWARA Y.

PA WPI; 1995-182987/24.

PT Novel anti-idiotype antibody against an human anticancer monoclonal  
antibody - and DNA sequences encoding the antibody, useful in  
pharmacology, medicine and biochemical fields.

PS Claim 11; Page 4; 28pp; Japanese.

CC A new anti-idiotype antibody against a human anticancer monoclonal  
CC antibody is claimed. This antibody contains in its heavy chain 3  
CC complementarity determining regions CDR1 (AAK74929-R74931), CDR2  
CC (AAK74932-R74935) and CDR3 (AAK74936-R74939). This is also true of the  
CC light chain which has its own CDR1 (AAK74944-R74946 and AAK85774), CDR2  
CC (AAK74947-R74949) and CDR3 (AAK74950-R74954). The antibody and DNA  
CC encoding it are useful in pharmacological, medical and biochemical  
CC fields.

SO Sequence 11 AA;

Query Match 90.0%; Score 45; DB 16; Length 11;  
Best Local Similarity 90.9%; Pred. No. 0.047;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11  
| | | | | | | | | |  
DB 1 kasqdvstava 11

#### RESULT 10

AAW62189 10  
ID AAW62189 standard; peptide; 11 AA.

AC AAW62189;

DT 21-SEP-1998 (first entry)

DE Mouse anti-HM1.24 antibody L chain V region CDR 1.

KW Mouse; human; humanised; anti-HM1.24 antibody; myeloma; FR; CDR;

KW framework region; complementarity determining region; antigenicity.  
 XX Mus sp.  
 XX MO9814580-A1.  
 PN 09-APR-1998.  
 XX 03-OCT-1997; 97WO-JP03553.  
 XX 04-OCT-1996; 96JP-0264756.  
 XX (CHUS ) CHUGAI SEIYAKU KK.  
 PA Koishihara Y, Kosaka M, Ohtomo T, Ono K, Tsuchiya M;  
 PI Yoshimura Y;  
 XX WPI: 1998-286421/25.  
 DR Humanised anti-HM1.24 antibody - for treatment of myeloma  
 XX  
 XX Claim 10; Page 102; 210pp; Japanese.  
 PS A humanised anti-HM1.24 antibody has been developed which comprises  
 CC human L and H chain C regions, and L and/or H chain V regions  
 CC containing material originating in mouse anti-HM1.24 antibody. The V  
 CC regions contain framework (FR) regions of human origin and  
 CC complementarity determining regions (CDR) of mouse origin, leading to  
 CC a reshaped humanised antibody. The C regions are human Ck (L-chain) and  
 CC human C gamma (especially C gamma 1) (H-chain). The FR regions of the  
 CC L chain V region are derived from human subtype HSG1 (e.g. from human  
 CC antibody RE1) and the FR regions of the H chain V region are derived  
 CC from human subtype HSG1 (e.g. FR1-3 from human antibody HG3 and FR4  
 CC from human antibody JH6). The present sequence represents mouse L chain  
 CC V region CDR 1 from the present invention. The antibodies are used for  
 CC the treatment of myeloma, especially by injection, intravenously,  
 CC intramuscularly or subcutaneously. The antibodies are used at 0.01-1000  
 CC (especially 5-100) mg/kg body weight. The humanised antibody has low  
 CC antigenicity and is therefore effective therapeutically in humans.  
 CC  
 XX Sequence 11 AA:  
 SQ  
 QY 1 KASQDVTTAVA 11  
 DB 1 kasqdvntava 11  
 Query Match 90.0%; Score 45; DB 19; Length 11;  
 Best Local Similarity 90.9%; Pred No. 0.047; 1; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 11  
 AAM57577  
 ID AAM57577 standard; peptide; 11 AA.  
 XX  
 XX AAM57577;  
 AC  
 XX 03-SEP-1998 (first entry)  
 DT  
 XX Chimeric L chain V region CDR-1 for an antibody against hPTRP.  
 DE  
 XX Chimeric; antibody; human parathormone related peptide; hPTRP; mouse;  
 KW L chain; H chain; hypercalcaemia; cancer; malignant lymphoma; CDR;  
 KW hypophosphataemia; pathogen; vitamin D resistance; V region; C region;  
 KW humanised.  
 XX  
 XX Synthetic.  
 OS Chimeric - Mus sp.  
 OS Chimeric - Homo sapiens.  
 XX  
 XX MO9813388-A1.

PD 02-APR-1998.  
 XX  
 XX 24-SEP-1997; 97WO-JP03382.  
 XX  
 XX 24-JUL-1997; 97JP-0214168.  
 PR 26-SEP-1996; 96JP-0255196.  
 XX  
 XX (CHUS ) CHUGAI SEIYAKU KK.  
 PA Sato K, Wakahara Y, Yabuta N;  
 PI WPI: 1998-230640/20.  
 DR  
 XX  
 XX New chimeric antibodies against human parathormone related  
 PT peptide(s) - useful for, e.g. treatment of hypercalcaemia and other  
 PR disorders caused by malignant neoplasms(s)  
 XX  
 XX Claim 9; Page 123; 182pp; Japanese.  
 PS  
 XX  
 XX New antibodies have been developed which are specific for human  
 CC parathormone related peptides (hPTRP). The antibodies comprise chimeric  
 CC L and/or H chains, where the C region is of human and L region of mouse.  
 CC origin. The present sequence represents a specifically claimed region of  
 CC an antibody of the invention. Host cells, transformed with vectors  
 CC containing DNA encoding antibodies of the invention, can be used to  
 CC produce the antibodies. The antibodies may be used to treat  
 CC hypercalcaemia, especially that due to a malignancy, e.g. cancers of  
 CC pancreas, lung, throat, larynx, tongue, gum, oesophagus, stomach, liver,  
 CC breast, kidney, bladder, womb or prostate or malignant lymphoma. They  
 CC may also be used for treatment of hypophosphataemia such as that due to  
 CC pathogens or to vitamin D resistance.  
 CC  
 XX Sequence 11 AA:  
 SQ  
 QY 1 KASQDVTTAVA 11  
 DB 1 kasqdvntava 11  
 Query Match 90.0%; Score 45; DB 19; Length 11;  
 Best Local Similarity 90.9%; Pred No. 0.047; 1; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 12  
 AAY02547  
 ID AAY02547 standard; Peptide; 11 AA.  
 XX  
 XX AAY02547;  
 AC  
 XX 16-JUL-1999 (first entry)  
 DT  
 XX Artificial CDR(1) of L chain V region of antiHM1.24 antibody.  
 DE  
 XX Reconstituted human antibody; peptide antigen HM1.24; framework region;  
 KW complementary determining region; CDR; anti-HM1.24 antibody; myeloma.  
 KW  
 XX Synthetic.  
 OS  
 XX MO9918212-A1.  
 PN 15-APR-1999.  
 PD 02-OCT-1998; 98WO-JP04469.  
 PF 03-OCT-1997; 97JP-0271726.  
 PR  
 XX (CHUS ) CHUGAI SEIYAKU KK.  
 PA Tsuchiya M;  
 PI WPI: 1999-277273/23.  
 DR





DT 13-NOV-2001 (first entry)  
 DE Amino acid sequence of a human peptide.  
 XX  
 KW Tissue decomposition inhibitor; parathyroid hormone; cancer cachexia;  
 KW septicemia; injury; muscular dystrophy; cytokine; interleukin-6;  
 KW granulocyte colony stimulating factor; interleukin-11;  
 KW leukemia inhibitory factor; weight loss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164249-A1.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 30-AUG-2000; 2000WO-JP05886.  
 XX  
 PR 28-FEB-2000; 2000JP-0052414.  
 XX  
 PA (CHUS ) CHUGAI SEIYAKU KK.  
 XX  
 PI Saito H, Tsunenari T, Onuma E, Sato K;  
 XX  
 DR WPI; 2001-550131/61.  
 XX  
 PT Tissue decomposition inhibitor that prevents parathyroid hormone  
 PT associated proteins from binding to its receptor  
 XX  
 PS Disclosure: Page 108; 132pp; Japanese.  
 XX  
 CC The specification describes a tissue decomposition inhibitor, which  
 CC comprises a substance that inhibits peptides associated with  
 CC parathyroid hormone (PTH) from binding with their receptor. The method  
 CC is used to inhibit tissue decomposition caused by cancer cachexia,  
 CC septicemia, heavy external injury or muscular dystrophy, and for  
 CC treating patients with elevated cytokine (interleukin-6, granulocyte  
 CC colony stimulating factor, interleukin-11 and leukemia inhibitory  
 CC factor) levels. It may also be used for preventing weight loss caused  
 CC by cancer cachexia. The present sequence represents a peptide, which is  
 CC used in the course of the invention.  
 XX  
 SQ Sequence 11 AA:

Query Match 90.0%; Score 45; DB 22; Length 11;  
 Best Local Similarity 90.9%; Pred. No. 0.047;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 KASQDVTTAVA 11  
 |||||  
 DB 1 kasqdvttava 11

Search completed: August 20, 2002, 11:34:38  
 Job time: 632 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:28:41 : Search time 25.74 Seconds  
(Without alignments)  
41.064 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_24\_34  
Perfect score: 50  
Sequence: 1 KASQDVTAVA 11

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	94.0	152	2 S30751	Ig kappa chain pre
2	46	92.0	93	2 S38559	Ig light chain V r
3	46	92.0	107	2 S32191	Ig kappa chain V r
4	45	92.0	107	2 S32192	Ig kappa chain V r
5	45	90.0	107	2 S09967	Ig kappa chain V-J
6	43	86.0	117	2 S42466	Ig kappa chain V r
7	42	84.0	119	2 P00265	Ig kappa chain V r
8	42	84.0	149	1 KYMS11	Ig kappa chain pre
9	41	82.0	106	2 PL0088	Ig kappa chain V r
10	40	80.0	98	2 PH1073	Ig light chain V r
11	39	78.0	94	2 F33730	Ig kappa chain V r
12	38	76.0	82	2 S38560	Ig light chain V r
13	38	76.0	90	2 S38561	Ig light chain V r
14	38	76.0	108	2 PL0083	Ig kappa chain V r
15	38	76.0	131	2 PL0207	anti-idiotypic ant
16	37	74.0	100	2 H38601	Ig kappa chain V r
17	36	72.0	1449	2 T30857	glucosyltransferas
18	36	72.0	1449	2 T30552	glucosyltransferas
19	34	68.0	45	2 B43859	ORF overlapping ca
20	34	68.0	108	2 PL0204	anti-DNA autoantib
21	34	68.0	328	2 AB1781	hypothetical prote
22	34	68.0	328	2 AC1405	hypothetical prote
23	34	68.0	374	2 AC0469	probable uroporphyr
24	34	68.0	389	2 AF0920	uroporphyrinogen I
25	34	68.0	393	2 S02185	uroporphyrin-III C
26	34	68.0	395	2 T21679	hypothetical prote
27	34	68.0	399	2 E91220	uroporphyrinogen I
28	34	68.0	399	2 G86066	uroporphyrinogen I
29	34	68.0	435	2 H87516	D-alanyl-D-alanine

30	34	68.0	501	2 F98050	lysozyme (EC 3.2.1
31	34	68.0	633	2 E31265	lipa protein - Shi
32	34	68.0	815	2 T00546	serine/threonine-s
33	33	66.0	100	2 T50611	hypothetical prote
34	33	66.0	226	2 T29210	hypothetical prote
35	33	66.0	485	1 B45343	glycoprotein gp13
36	33	66.0	485	2 T42559	envelope protein 1
37	33	66.0	822	2 T25866	hypothetical prote
38	33	66.0	2396	2 T13714	kakapo gene protei
39	32	64.0	165	2 JC5737	membrane-bound pro
40	32	64.0	191	2 AG2916	conserved hypochet
41	32	64.0	235	2 B97691	pHr protein (Af02
42	32	64.0	263	2 D83085	conserved hypochet
43	32	64.0	330	2 T41917	hypothetical prote
44	32	64.0	371	2 E84709	hypothetical prote
45	32	64.0	398	2 T26274	hypothetical prote

## ALIGNMENTS

RESULT 1  
S30751  
Ig kappa chain precursor V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 21-Jan-2000  
C:Accession: S30751  
R:Grant, F.J.; Levin, S.D.; Gilbert, T.; Kindsvogel, W.  
Nucleic Acids Res. 15, 5496, 1987  
A:Title: Improved RNA sequencing method to determine immunoglobulin mRNA sequence.  
A:Reference number: S30751; MUID:87260030  
A:Accession: S30751  
A:Molecule type: mRNA  
A:Residues: 1-152 <GBR>  
A:Cross-references: EMBL:X05877; NID:952195; PIDN:CAA29301.1; PID:952196  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:45-119/Domain: immunoglobulin homology <IMM>  
F:138-152/Domain: C region (C-kappa) (fragment) #status predicted <CRE>

Query Match 94.0%; Score 47; DB 2; Length 152;  
Best Local Similarity 90.9%; Pred. No. 0.042;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTAVA 11  
|||||  
Db 53 KASQDVTAVS 63

RESULT 2  
S38559  
Ig light chain V region (ASWAI) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jan-2000  
R:Monestier, M.; Losman, L.J.; Novick, K.E.; Aris, J.P.  
Submitted to the EMBL Data Library, September 1993  
A:Description: Molecular analysis of mercury-induced anti-nuclear antibodies in R-  
A:Reference number: S38559  
A:Accession: S38559  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-93 <MON>  
A:Cross-references: EMBL:X75101; NID:9414145; PIDN:CAA52992.1; PID:9414146  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: immunoglobulin  
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 92.0%; Score 46; DB 2; Length 93;  
Best Local Similarity 90.9%; Pred. No. 0.04;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11  
|||||:||||  
Db 24 KASQDVSTAVA 34

## RESULT 3

S32191  
Ig kappa chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 21-Jan-2000  
C:Accession: S32191  
R:Izui, S.  
Submitted to the EMBL Data Library, February 1993  
A:Reference number: S32185  
A:Accession: S32191  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-107 <I2U>  
A:Cross-references: EMBL:X70095; NID:9288260; PIDN:CAA49700.1; PID:9288261  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 92.0%; Score 46; DB 2; Length 107;  
Best Local Similarity 90.9%; Pred. No. 0.047;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11  
|||||:||||  
Db 24 KASQDVSTAVA 34

## RESULT 4

S32192  
Ig kappa chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 21-Jan-2000  
C:Accession: S32192  
R:Izui, S.  
Submitted to the EMBL Data Library, February 1993  
A:Reference number: S32185  
A:Accession: S32192  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-107 <I2U>  
A:Cross-references: EMBL:X70097; NID:9288262; PIDN:CAA49701.1; PID:9288263  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 92.0%; Score 46; DB 2; Length 107;  
Best Local Similarity 90.9%; Pred. No. 0.047;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11  
|||||:||||  
Db 24 KASQDVSTAVA 34

## RESULT 5

S09967  
Ig kappa chain V-J region (31-9D) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 16-Aug-1996  
C:Accession: S09967  
R:Reininger, L.; Shibata, T.; Ozaki, S.; Shirai, T.; Jaton, J.C.; Izui, S.  
Eur. J. Immunol. 20, 771-777, 1990  
A:Title: Variable region sequences of pathogenic anti-mouse red blood cell autoantibodies  
A:Reference number: S09955; MUID:9026328  
A:Accession: S09967

A:Molecule type: mRNA  
A:Residues: 1-107 <REI>  
A:Cross-references: EMBL:X51855  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 90.0%; Score 45; DB 2; Length 107;  
Best Local Similarity 90.9%; Pred. No. 0.074;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11  
|||||:||||  
Db 24 KASQDVNTAVA 34

## RESULT 6

S42466  
Ig kappa chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jan-2000  
C:Accession: S42466  
R:Shyanov, P.A.; Bessalov, I.A.; Terletskaya, H.N.; Deyev, S.M.  
Submitted to the EMBL Data Library, March 1994  
A:Reference number: S42466  
A:Accession: S42466  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-117 <SHI>  
A:Cross-references: EMBL:X78108; NID:9460824; PIDN:CAA54998.1; PID:9460825  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:26-100/Domain: immunoglobulin homology <IMM>

Query Match 86.0%; Score 43; DB 2; Length 117;  
Best Local Similarity 90.9%; Pred. No. 0.21;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11  
|||||:||||  
Db 34 KASQDVGTAVA 44

## RESULT 7

PQ0265  
Ig kappa chain V region (MC1) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 21-Jan-2000  
C:Accession: PQ0265  
R:Lothman, K.L.; Carrillo, M.A.; Kennedy, R.C.  
Gene 105, 283-284, 1991  
A:Title: Sequence analysis of the variable region of a mouse gene encoding a monoclonal antibody  
A:Reference number: PQ0265; MUID:92039046  
A:Accession: PQ0265  
A:Molecule type: mRNA  
A:Residues: 1-119 <LOH>  
A:Cross-references: GB:M59985  
C:Comment: This protein recognizes a restricted idiotype associated with antibodies s  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:28-102/Domain: immunoglobulin homology <IMM>  
F:36-46/Region: complementarity-determining 1  
F:62-68/Region: complementarity-determining 2  
F:101-109/Region: complementarity-determining 3

Query Match 84.0%; Score 42; DB 2; Length 119;  
Best Local Similarity 81.8%; Pred. No. 0.34;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASQDVTTAVA 11  
|||||:||||

Db 36 KASQDVSTTVA 46

RESULT 8

KMS11

Ig kappa chain precursor V region (MPC11) - mouse  
C:Species: Mus musculus (house mouse)

C:Date: 19-Feb-1984 #sequence\_revision 19-Feb-1984 #text\_change 21-Jan-2000

C:Accession: A90823; A90753; A90298; A01916

R:Kelley, D.E.; Coleclough, C.; Perry, R.P.

Cell 29, 681-689, 1982

A:Title: Functional significance and evolutionary development of the 5'-terminal regions

A:Reference number: A90823; MWID:83001944

A:Accession: A90823

A:Molecule type: DNA

A:Residues: 1-71 &lt;KRL&gt;

A&gt;Note: the sequence was determined from the differentiated gene

R:Rabbits, T.H.; Hamlyn, P.H.; Mathysens, G.; Roe, B.A.

Can. J. Biochem. 58, 176-187, 1980

A:Title: The variability, arrangement, and rearrangement of immunoglobulin genes.

A:Reference number: A90753; MWID:80176554

A:Accession: A90753

A:Molecule type: mRNA

A:Residues: 41-149 &lt;RAB&gt;

R:Smith, G.P.

Biochem. J. 171, 337-347, 1978

A:Title: Sequence of the full-length immunoglobulin kappa-chain of mouse myeloma MPC 11.

A:Reference number: A90298; MWID:78186617

A:Accession: A90298

A:Contents: myeloma protein MPC11

A:Molecule type: protein

A:Residues: 30-149 &lt;SMT&gt;

A&gt;Note: the amidation states of residues 58, 69, 101, 111, 120, 122, and 123 were not determined

C:Comment: The mature chain has 12 additional residues at its amino end, due to a tandem

42 corresponds to the amino-terminal residue of typical kappa chains.

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kappa)

chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into IgA

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: duplication; heterotetramer; immunoglobulin

F:1-29/Domain: signal sequence #status predicted &lt;Sig&gt;

F:30-149/Product: Ig kappa chain V region (MPC11) #status experimental &lt;Mat&gt;

F:57-131/Domain: immunoglobulin homology &lt;IMM&gt;

F:64-129/Disulfide bonds: #status predicted

Query Match

Best Local Similarity

Matches 9; Conservative

1; Mismatches

1; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

Query Match 82.0%; Score 41; DB 2; Length 106;  
Best Local Similarity 81.8%; Pred. No. 0.47;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;OY 1 KASQDVTTAVA 11  
Db 24 KASQDVNSAVA 34

RESULT 10

PHI073

Ig light chain V region (clone 17s.130) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jan-2000

C:Accession: PHI073

R:Thilman, D.M.; Jou, N.T.; Hill, R.J.; Marlon, T.N.

J. Exp. Med. 176, 761-779, 1992

A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective

A:Reference number: PH0971; MWID:92381444

A:Accession: PHI073

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-98 &lt;TIL&gt;

A:Experimental source: B cell, strain [NZB x NZW]F1

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: immunoglobulin

F:16-90/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match

Best Local Similarity

Matches 9; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

Query Match 78.0%; Score 39; DB 2; Length 94;  
Best Local Similarity 81.8%; Pred. No. 1.1;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;OY 1 KASQDVTTAVA 11  
Db 24 KASQDVNSAVA 34

RESULT 12

S38560

Ig light chain V region (ASWB1) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999

C:Accession: S38560

R:Lawler, A.M.; Kearney, J.F.; Kuehl, M.; Gearhart, P.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 6744-6747, 1989

A:Title: Early rearrangements of genes encoding murine immunoglobulin kappa-chains, u

A:Reference number: A33730; MWID:89367325

A:Accession: F33730

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-94 &lt;LAW&gt;

A:Cross-references: GB:M26002; MID:q197123; PIDN:AAA38918.1; PID:q554119

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:16-90/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match

Best Local Similarity

Matches 9; Conservative

1; Mismatches

1; Indels

0; Gaps

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

0;

C;Accession: S38560  
R;Monestier, M.; Losman, L.J.; Novick, R.E.; Aris, J.P.  
submitted to the EMBL Data Library, September 1993  
A;Description: Molecular analysis of mercury-induced anti-nuclear antibodies in H-2s M  
A;Reference number: S38559  
A;Accession: S38560  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-82 <MON>  
A;Cross-references: EMBL:X75102; NID:9414147; PIDN:CAA52993.1; PID:9414148  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: immunoglobulin

Query Match 76.0%; Score 38; DB 2; Length 82;  
Best Local Similarity 81.8%; Pred. No. 1.5;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KASQDVTTAVA 11  
||||:|||||  
DB 24 KASQNVGTAVA 34

RESULT 13  
S38561  
Ig light chain V region (ASMP1) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jan-2000  
C;Accession: S38561  
R;Monestier, M.; Losman, L.J.; Novick, R.E.; Aris, J.P.  
submitted to the EMBL Data Library, September 1993  
A;Description: Molecular analysis of mercury-induced anti-nuclear antibodies in H-2s M  
A;Reference number: S38559  
A;Accession: S38561  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-90 <MON>  
A;Cross-references: EMBL:X75103; NID:9414155; PIDN:CAA52994.1; PID:9414156  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: immunoglobulin  
F;16-90/Domain: immunoglobulin homology <IMM>

Query Match 76.0%; Score 38; DB 2; Length 90;  
Best Local Similarity 81.8%; Pred. No. 1.6;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KASQDVTTAVA 11  
||||:|||||  
DB 24 KASQNVGTAVA 34

RESULT 14  
PL0083  
Ig kappa chain V region (E3) - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 21-Jan-2000  
C;Accession: PL0083  
R;Meek, K.; Hasemann, C.; Pollok, B.; Alkan, S.S.; Bralt, M.; Slaoui, M.; Urbain, J.; Ca  
J. Exp. Med. 169, 519-533, 1989  
A;Title: Structural characterization of antidiotytic antibodies: evidence that Ab2s are  
A;Reference number: PL0080; MUID:89094248  
A;Accession: PL0083  
A;Molecule type: mRNA  
A;Residues: 1-108 <MEE>  
A;Experimental source: strain BALB/c  
A;Note: the sequence shown here is from the V kappa region of an antidiotytic monoclonal  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F;16-91/Domain: immunoglobulin homology <IMM>

Query Match 76.0%; Score 38; DB 2; Length 108;

Best Local Similarity 81.8%; Pred. No. 2;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KASQDVTTAVA 11  
||||:|||||  
DB 24 KASQNVGTAVA 34

RESULT 15  
PL0207  
anti-idiotypic antibody E225, kappa chain V region precursor - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 21-Jan-2000  
C;Accession: PL0207  
R;Souchon, H.; Doyen, N.; Riotelet, M.M.; Rougeon, F.; Poljak, R.J.  
Mol. Immunol. 27, 429-433, 1990  
A;Title: Nucleotide sequence of the VH, VL regions of an anti-idiotypic antibody reac  
A;Reference number: PL0207; MUID:90309764  
A;Accession: PL0207  
A;Molecule type: mRNA  
A;Residues: 1-131 <SGU>  
A;Note: This sequence corresponds to subgroup V mouse immunoglobulin light chain  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;36-110/Domain: immunoglobulin homology <IMM>  
F;44-54/Region: complementarity-determining 1  
F;70-76/Region: complementarity-determining 2  
F;109-117/Region: complementarity-determining 3  
F;116-127/Region: J region

Query Match 76.0%; Score 38; DB 2; Length 131;  
Best Local Similarity 81.8%; Pred. No. 2.4;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KASQDVTTAVA 11  
||||:|||||  
DB 44 KASQDVRTAVA 54

Search completed: August 20, 2002, 11:35:17  
Job time: 396 sec

100-443888-1000

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:34:41 ; Search time 15.95 Seconds

(Without alignments)  
26.703 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_24\_34  
Perfect score: 50  
Sequence: 1 KASQDVTAVV 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	84.0	149	KV5A_MOUSE	P01633 mus musculi
2	34	68.0	393	HEMX_ECOLI	P09127 escherichia
3	34	68.0	633	IPAA_SHIFL	P18010 shigella fl
4	33	66.0	485	VGIC_HSYE4	P22596 equine herp
5	33	66.0	975	CSEL_DROME	P09201 drosophila
6	32	64.0	165	ATPF_STROM	P35765 streptococc
7	32	64.0	474	FLA_LEGMI	P33606 legionella
8	32	64.0	481	HTR3_HALN1	P09400 halobacteri
9	32	64.0	481	HTR3_HALSA	P08362 halobacteri
10	31	62.0	108	KV1Y_HUMAN	P08362 homo sapien
11	31	62.0	279	GV12_HALN1	P09400 halobacteri
12	31	62.0	549	TCPA_CAEEL	P11988 caenorhabdi
13	31	62.0	788	ACON_CAEEL	P44455 caenorhabdi
14	31	62.0	1041	EGT2_YEAST	P24835 saccharomyc
15	31	62.0	1275	YAU9_SCHPO	P48946 synecocyst
16	30	60.0	71	RS18_SYNY3	P00607 candida tro
17	30	60.0	160	VATL_CANTR	P41847 caenorhabdi
18	30	60.0	238	Y097_CAEEL	P39039 mus musculi
19	30	60.0	239	MABA_MOUSE	P24076 medicago sa
20	30	60.0	325	GBLP_MEDSA	P03986 glycyne max
21	30	60.0	325	GBLP_SOYBN	P00767 enterococcu
22	30	60.0	442	FTSA_ENTHR	P08365 escherichia
23	30	60.0	450	CRPD_ECOLI	P09127 escherichia
24	30	60.0	459	PEDA_LACLA	P12889 equine herp
25	30	60.0	468	VGIC_HSYE4	P22596 equine herp
26	30	60.0	483	NDAD_ALICX	P35887 caulobacter
27	30	60.0	490	DNAA_CAUCR	P44212 alcaligenes
28	30	60.0	498	NDAD_ALICX	P35887 caulobacter
29	30	60.0	562	CHIT_YEAST	P29029 saccharomyc
30	30	60.0	750	LEU2_RHINI	P55811 rhizopus ni
31	30	60.0	793	NETB_DROME	P24566 drosophila
32	30	60.0	901	VGLB_GPCMV	P09024 guinea pig
33	30	60.0	969	AGLU_SCHPO	P09024 schizosacch

34	30	60.0	984	1	SV13_MOUSE	Q04891 mus musculi
35	30	60.0	2334	1	WAPA_BACSV	Q07833 bacillus su
36	30	60.0	2512	1	POLN_SINDV	P03317 sindbis vir
37	30	60.0	2514	30	POLN_SINDV	P27283 sindbis vir
38	29	58.0	81	1	Y039_STMP7	P27368 synecococc
39	29	58.0	107	1	FER_PSLA	P24806 psalteriomo
40	29	58.0	145	1	LIRP_LOOMI	P15131 locusta mig
41	29	58.0	181	1	RBS1_BRANA	P05346 brassica na
42	29	58.0	204	1	PS2L_ARATH	Q23714 arabidopsis
43	29	58.0	222	1	SODM_MOUSE	P06671 mus musculi
44	29	58.0	222	1	SODM_RAT	P07895 rattus norv
45	29	58.0	247	1	143B_SOYBN	Q96451 glycine max

## ALIGNMENTS

```

RESULT 1
KV5A_MOUSE STANDARD: PRT; 149 AA.
ID KV5A_MOUSE
AC P01633:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-V region MPC11 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=83001944; PubMed=6288267;
RA Kelley D.E., Coleclough C., Perry R.P.;
RT "Functional significance and evolutionary development of the
RT 5'-terminal regions of immunoglobulin variable-region genes.";
RL Cell 29:681-689(1982).
RN [2]
RP SEQUENCE OF 41-149 FROM N.A.
RX MEDLINE=80176554; PubMed=6245773;
RA Rabbits T.H., Hamlyn P.H., Matlyssens G., Roe B.A.;
RT "The variability, arrangement, and rearrangement of immunoglobulin
RT genes.";
RL Can. J. Biochem. 58:176-187(1980).
RN [3]
RP SEQUENCE OF 30-149.
RX MEDLINE=78186617; PubMed=418775;
RA Smith G.P.;
RT "Sequence of the full-length immunoglobulin kappa-chain of mouse
RT myeloma MPC 11.";
RL Biochem. J. 171:337-347(1978).
CC -!- MISCELLANEOUS: THE MATURE CHAIN HAS 12 ADDITIONAL RESIDUES AT ITS
CC AMINO END, DUE TO A TANDEM DUPLICATION OF 36 NUCLEOTIDES AFTER THE
CC CODON FOR RESIDUE 36. RESIDUE 42 CORRESPONDS TO THE AMINO-TERMINAL
CC RESIDUE OF TYPICAL KAPPA CHAINS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, J00561, AAA8776.1; -.
CC PIR, A01916; KWSM11.
CC InterPro, IPR003006; Ig_MHC.
CC InterPro, IPR003596; Ig_V.
CC Pfam, PF00047; Ig; 1.
CC SMART, SM00406; IGV; 1.
CC Immunoglobulin V region; Signal; Repeat.
CC CHAIN 1
CC SIGNAL 29
CC DOMAIN 42 64
CC FRAMEWORK-1.

```

```

FT DOMAIN 65 75 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 76 90 FRAMEWORK-2.
FT DOMAIN 91 97 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 98 129 FRAMEWORK-3.
FT DOMAIN 130 138 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 139 148 FRAMEWORK-4.
FT REPEAT 26 35
FT REPEAT 38 47
FT NON_TER 149 149
SQ SEQUENCE 149 AA; 16434 MW; B0480C87B682AC3E CRC64;

Query Match 84.0%; Score 42; DB 1; Length 149;
Best Local Similarity 81.8%; Pred. No. 0.14;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Cy 1 KASODVTTAVA 11
    |||||:|
Db 65 KASODVSTTVA 75

RESULT 2
HEMX_ECOLI STANDARD; PRT: 393 AA.
ID P09127;
AC 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative uroporphyrin-III C-methyltransferase (EC 2.1.1.107) (Urogen
DE III methylase) (ORF X).
GN HMXX OR B3803.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=89098348; PubMed=3062586;
RA Sasarman A., Echelard Y., Letowski J., Tardif D., Drolet M.;
RT "Nucleotide sequence of the hemx gene, the third member of the uro
RT operon of Escherichia coli K12."
RL Nucleic Acids Res. 16:11835-11835(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / CS520;
RX MEDLINE=89041586; PubMed=3054815;
RA Alefounder P.R.;
RT "The sequence of hemc, hemd and two additional E. coli genes."
RL Nucleic Acids Res. 16:9871-9871(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes."
RL Science 257:771-778(1992).
RN [4]
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12."
RL Electrophoresis 18:1259-1313(1997).
CC -1- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + uroporphyrin III =
CC 2 S-adenosyl-L-homocysteine + 5-hydroxytryptophan.
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF SIROHEME AND COBALAMIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13406; CAA31772.1; -
DR EMBL; X12614; CAA31134.1; -
DR EMBL; M87049; AAA67599.1; -
DR EMBL; AE000456; AAC76806.1; -
DR PIR; S02185; S02185.
DR PIR; S30693; S30693.
DR EC02DBASE; B043.0; 6TH EDITION.
DR EC02DBASE; B043.1; 6TH EDITION.
DR EcGene; EG10433; hemx.
KW Porphyrin biosynthesis; Transferase; Methyltransferase;
KW complete proteome.
SQ SEQUENCE 393 AA; 42963 MW; 9D272C6401DDE354 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 393;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Cy 2 ASODVTTAVA 11
    |||||:|
Db 151 SDODVTTAA 160

RESULT 3
IPAA_SHIFL STANDARD; PRT: 633 AA.
ID P18010;
AC 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 70 kDa antigen.
GN IPAA.
OS Shigella flexneri.
OG Plasmid 210 kb invasion pWR100.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_Taxid=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M90T / SEROTYPE 5;
RX MEDLINE=90221912; PubMed=2183200;
RA Venkatesan M.M., Buysse J.M.;
RT "Nucleotide sequence of invasion plasmid antigen gene ipaA from
RT Shigella flexneri 5."
RL Nucleic Acids Res. 18:1648-1648(1990).
RN [2]
RP SEQUENCE OF 1-88 FROM N.A.
RC STRAIN-M90T / SEROTYPE 5;
RX MEDLINE=89057927; PubMed=3057506;
RA Venkatesan M.M., Buysse J.M., Kopecko D.J.;
RT "Characterization of invasion plasmid antigen genes (ipaBCD) from
RT Shigella flexneri."
RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).
CC -1- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
CC EPITHELIAL CELLS.
CC -----
CC MISCELLANEOUS: SYNTHESIS OF THIS IMMUNOGEN IS RERESSED AT 30
CC DEGREES CELSIUS AND RESTORED AT 37 DEGREES CELSIUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17628; CAA35624.1; -
DR EMBL; J04117; AAA26525.1; -

```

DR PIR: E31265; E31265.  
DR PIR: S12763; S12763.  
KW Antigen: Plasmid: Virulence.  
SQ SEQUENCE 633 AA; 70093 MW; 2F804F45355E4751 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 633;  
Best Local Similarity 54.5%; Pred. No. 27;  
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQVTTAV 11  
DB 614 KAAQVTTSL 624

RESULT 4  
VGLC\_HSV4 STANDARD; PRT; 485 AA.  
AC P22596;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Glycoprotein C precursor (Glycoprotein 13).  
GN GC OR GP13.  
OS Equine herpesvirus type 4 (strain 1942) (EHV-4) (Equine herpesvirus  
OS type 1 subtype 2).  
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicelloviruses.  
OX NCBI\_TaxID=10333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91021040; PubMed=2171212;  
RA Nicolson L., Onions D.E.;  
RT "The nucleotide sequence of the equine herpesvirus 4 gc gene  
RT homologue";  
RL Virology 179:378-387(1990).  
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M58031; AAA46083.1; -  
DR EMBL: A21044; CA01528.1; -  
DR PIR: B45343; B45343.  
DR InterPro: IPR001038; EHV\_gp13.  
DR InterPro: IPR001654; Marek\_A.  
DR Pfam: PF02124; Marek\_A.1.  
DR PRINTS: PR00668; GLYCOPROTEIN C.  
KW Glycoprotein; Transmembrane; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 485 GLYCOPROTEIN C.  
FT DOMAIN 33 444 EXTRACELLULAR.  
FT TRANSMEM 445 468  
FT CARBOHYD 60 60  
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 485 AA; 52509 MW; 599AC10B9C2598E CRC64;

Query Match 66.0%; Score 33; DB 1; Length 485;

Best Local Similarity 87.5%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 SQVTTAV 10  
DB 382 SQVTTGV 389

RESULT 5  
CSEL\_DROME STANDARD; PRT; 975 AA.  
AC Q9XZU1; Q9VJH4; Q9UB14;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Importin-alpha re-exporter (Cellular apoptosis susceptibility protein  
DE homolog).  
GN CAS OR BCDNA:ID14270 OR CG13281.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tekotte H., Davis I.;  
RT "Drosophila melanogaster CAS";  
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Prankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreffi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weissbrock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Tsang G., Brokstein P., Friese E., Harvey D., Evans-Holm M.,

```

RA Lewis S.E., Suh C., Rubin G.M.;
RT "Full length Drosophila melanogaster cDNA sequence.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC
CC -1- FUNCTION: Export receptor for importin alpha. Mediates importin-
CC alpha reexport from the nucleus to the cytoplasm after import
CC substrates have been released into the nucleoplasm (By
CC similarity).
CC -1- SUBUNIT: Binds with high affinity to importin-alpha only in the
CC presence of RanGTP (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CSE1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ238857; CAB42967.1; -.
DR EMBL: AE003652; AAF53575.1; -.
DR EMBL: AF132562; AAD27861.1; -.
DR FlyBase: FBgn0022213; Cas.
DR InterPro: IPR001494; IBN_NT.
KW Transport; Protein transport; Nuclear protein.
FT CONFLICT 194 194 K -> T (IN REF. 2).
FT CONFLICT 331 331 Q -> P (IN REF. 1).
FT CONFLICT 589 590 AA -> GR (IN REF. 1).
SQ SEQUENCE 975 AA; 110168 MW; 5C9BAllC08D49C11 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 975;
Best Local Similarity 63.6%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 KASQDVTAV 11
Db 52 KASQDVTAV 62

RESULT 6
ATPE_STRMU STANDARD; PRT; 165 AA.
AC P95785;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP synthase B chain (EC 3.6.3.14).
GN ATPF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCB1_TaxID=1309;
RX STRAIN=GS-5; Pubmed=8996091;
RX MEDLINE=97149283; Pubmed=8996091;
RA Smith A.J., Quivey R.G., Faustoferrl R.C.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus mutans
RT membrane-bound, proton-translocating ATPase operon.";
RL Gene 183:87-96(1996)

-1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
HAS THREE MAIN SUBUNITS: A, B AND C.
-1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31170; AAD13379.1; -.
DR InterPro: IPR002146; ATP-synt_B.
DR Pfam: PF00430; ATP-synt_B.1.
KW Hydrogen ion transport; Transmembrane; CF(0).
FT TRANSMEM 7 27 POTENTIAL.
SQ SEQUENCE 165 AA; 17952 MW; BIDAC1206386113 CRC64;

Query Match 64.0%; Score 32; DB 1; Length 165;
Best Local Similarity 54.5%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KASQDVTAV 11
Db 110 KASQDVTAV 120

RESULT 7
FLA_LEGMI STANDARD; PRT; 474 AA.
ID P53606;
AC P53606;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Flagellin.
GN FLA.
OS Legionella micdadei.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Tallochia.
ON NCB1_TaxID=451;
RX MEDLINE=96163446; Pubmed=8562027;
RX Bangborg J., Hindersson P., Shand G., Holby N.;
RT "The Legionella micdadei flagellin: expression in Escherichia coli K
RT 12 and DNA sequence of the gene.";
RL APMS 103:869-877(1995).
-1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
-1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X84699; CAA59172.1; -.
DR InterPro: IPR001029; Flagellin_C.
DR InterPro: IPR001492; Flagellin_N.
DR Pfam: PF00700; Flagellin_C.1.
DR Pfam: PF00669; Flagellin_N.1.
DR PRINTS: PR00207; FLAGELLIN.
DR Prodom: PD000316; Flagellin_C.1.
KW Flagella.
SQ SEQUENCE 474 AA; 48113 MW; 5C9DD5C8AEF9D07B CRC64;

Query Match 64.0%; Score 32; DB 1; Length 474;
Best Local Similarity 70.0%; Pred. No. 52;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 ASQDVTAV 11
Db 211 ASQDVTAV 220

RESULT 8

```

```

HTR3_HALN1
ID HTR3_HALN1 STANDARD: PRT: 481 AA.
AC 09H000:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Halobacterial transducer protein III.
GN HTR9 OR VNG1395G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Ballea N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weill R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angerme C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1."
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL
CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSOR DIMERIZATION DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE005058; ANG19717.1; -
DR InterPro: IPR004089; Chemotaxis_transducer.
DR InterPro: IPR003660; HAMp.
DR InterPro: IPR004090; Me_Chemotaxis.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00015; MCPsignal; 1.
DR PRINTS: PR00260; CHEMTRNSDUCR.
DR SMART: SM00304; HAMp; 1.
DR SMART: SM00283; MA; 1.
DR SMART: SM00091; PAS; 1.
DR PROSITE: PS50112; PAS; 1.
DR Transducer: Complete proteome.
KW Transducer: Complete proteome.
FT DOMAIN 10 81 PAS.
FT SEQUENCE 481 AA; 50773 MW; B33B55DBFE206D1 CRC64;
SQ

```

```

Query Match 64.0%; Score 32; DB 1; Length 481;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 3 S0DYTTAVA 11
: :|:|||||
DB 447 ADDITTAVA 455

```

```

RESULT 9
HTR3_HALSA
ID HTR3_HALSA STANDARD: PRT: 481 AA.
AC 048316;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Halobacterial transducer protein III.
GN HTR9 OR HTP111.

```

```

OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelt D.;
RA "A family of halobacterial transducer proteins."
RA FEMS Microbiol. Lett. 139:161-168(1996).
RL
CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSOR DIMERIZATION DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X95588; CA64840.1; -
DR HSSP: P02942; 1007.
DR InterPro: IPR004089; Chemotaxis_transducer.
DR InterPro: IPR003660; HAMp.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00015; MCPsignal; 1.
DR SMART: SM00304; HAMp; 1.
DR SMART: SM00283; MA; 1.
DR SMART: SM00091; PAS; 1.
DR PROSITE: PS50112; PAS; 1.
DR Transducer.
KW Transducer.
FT DOMAIN 10 81 PAS.
FT SEQUENCE 481 AA; 50728 MW; B283A873F7844A CRC64;
SQ

```

```

Query Match 64.0%; Score 32; DB 1; Length 481;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 3 S0DYTTAVA 11
: :|:|||||
DB 447 ADDITTAVA 455

```

```

RESULT 10
KV1Y_HUMAN
ID KV1Y_HUMAN STANDARD: PRT: 108 AA.
AC P80362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-1 region WAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE AND X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=95086080; PubMed=7993911;
RA Huang D.-B., Chang C.-H., Ainsworth C., Bruenger A.T., Eulitz M.,
RA Solomon A., Stevens F.J., Schiffer M.;
RA "Comparison of crystal structures of two homologous proteins:
RA structural origin of altered domain interactions in immunoglobulin
RA light-chain dimers."
RA Biochemistry 33:14848-14857(1994).
RL [2]
RN [2]
SEQUENCE OF 1-35.

```

RX MEDLINE-81267384; PubMed-6167731;  
 RA Stevens F.J., Westholm F.A., Panagiotopoulos N., Schiffer M.,  
 RA Popp R.A., Solomon A.;  
 RT "Characterization and preliminary crystallographic data on the VL-  
 RT related fragment of the human Ki Bence Jones protein Mat.";  
 RL J. Mol. Biol. 147:185-193(1981).  
 CC -1- MISCELLANEOUS: THIS IS A BENGE-JONES PROTEIN.  
 DR PDB: 1MTL; 01-NOV-94.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IGV; 1.  
 KM Immunoglobulin V region; Bence-Jones protein; 3D-structure.  
 FT DOMAIN 1 23  
 FT DOMAIN 24 34 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 35 49 FRAMEWORK-2.  
 FT DOMAIN 50 56 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 57 88 FRAMEWORK-3.  
 FT DOMAIN 89 97 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 98 107 FRAMEWORK-4.  
 FT DISULFID 23 88  
 FT CONFLICT 30 31 TN -> SD (IN REF. 2).  
 FT NON\_TER 108  
 SQ SEQUENCE 108 AA; 11737 MW; D9D941B3F0FAE697 CRC64;

Query Match 62.0%; Score 31; DB 1; Length 108;  
 Best Local Similarity 60.0%; Pred. No. 18;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KASODVTAV 10  
 Db 24 KASODITNYV 33

RESULT 11  
 ID GYL2\_HALN1 STANDARD; PRT; 279 AA.  
 AC Q9HHU0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GYPL protein 2.  
 GN GYPL2 OR VNG6229G.  
 OS Halobacterium sp. (strain NRC-1).  
 OS Plasmid pNRC200.  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 ON NCBI\_TaxID=64091;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20504483; PubMed-11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lesky S.R., Baliga N.S., Thorsson V., Shroga J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Meddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohnschoder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 RL -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS  
 VESICLE SYNTHESIS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AE005156; AAC20886.1; ALT\_INT.  
 KM Gas vesicle; Plasmid; Complete proteome.  
 SQ SEQUENCE 279 AA; 30376 MW; 543174A72EDE78B6 CRC64;  
 Query Match 62.0%; Score 31; DB 1; Length 279;  
 Best Local Similarity 66.7%; Pred. No. 49;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 ASODVTAV 10  
 Db 206 ASODTTTSL 214

RESULT 12  
 ID TCPA\_CAEEL STANDARD; PRT; 549 AA.  
 AC P41988; 022228;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE T-complex protein 1, alpha subunit (TCP-1-alpha) (CCR-alpha).  
 DE CCR-1 OR TCP-1 OR T05C12.7.  
 GN Caenorhabditis elegans.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 ON NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA MEDLINE-95278751; PubMed-7758963;  
 RA Leroux M.R., Candido E.P.M.;  
 RT "Molecular analysis of Caenorhabditis elegans tcp-1, a gene encoding  
 RT a chaperonin protein.";  
 RL Gene 156:241-246(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Burton J.;  
 RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON  
 CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF  
 CC ACTIN AND TUBULIN.  
 CC -1- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KDA THAT  
 CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U07941; AAB05072.1; -;  
 DR EMBL: Z66500; CAA91308.1; -;  
 DR HSSP: P48424; 1A6D.  
 DR WormPep: T05C12.7; CE02319.  
 DR InterPro: IPR002423; TCP1.cpn60.  
 DR InterPro: IPR002194; TCP\_1.  
 DR Pfam: PF00118; cpn60\_TCP1; 1.  
 DR PRINTS: PR00304; TCOMPLEXTCP1.  
 DR PROSITE: PS00750; TCP1\_1; 1.  
 DR PROSITE: PS00751; TCP1\_2; 1.  
 DR PROSITE: PS00995; TCP1\_3; 1.  
 KM Chaperone; ATP-binding; Multigene family.  
 FT CONFLICT 469 469 F -> K (IN REF. 2).  
 SQ SEQUENCE 549 AA; 58821 MW; D7C6D661FA391407 CRC64;

Query Match 62.0%; Score 31; DB 1; Length 549;

Best Local Similarity 77.8%; Pred. No. 97;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 3 SODVTAVA 11  
11:11 111  
Db 23 SODVTAVA 31

RESULT 13  
ACON.CAEEL STANDARD; PRT; 788 AA.  
ID ACON.CAEEL  
AC P34455;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Probable aconitate hydratase, mitochondrial precursor (EC 4.2.1.3)  
DE (Citrate hydro-lyase) (Aconitase).  
GN P54H12.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxId=6239;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Sams M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Stalston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohlsman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
CC -1- CATALYTIC ACTIVITY: Citrate = cis-aconitate + H(2)O.  
CC -1- COFACTOR: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-  
4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC  
CC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND ISOCITRATE (BY  
CC SIMILARITY).  
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Mitochondrial (by similarity).  
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/TPM ISOMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L25599; AAA28050.1; -  
DR PIR: S44831; S44831.  
DR HSP: P20004; 1AMJ.  
DR WormPep: F54H12.1; CE00516.  
DR InterPro: IPR001030; Aconitase.  
DR InterPro: IPR000573; Aconitase\_C.  
DR Pfam: PF00330; aconitase\_1.  
DR Pfam: PF00694; aconitase\_C; 1.  
DR PRINTS: PR00415; ACONITASE.  
DR PRODOM: PD000511; Aconitase; 1.  
DR PROSITE: PS00450; ACONITASE\_1; 1.  
DR PROSITE: PS01244; ACONITASE\_2; 1.  
KM Hypothetical protein, lyase, tricarboxylic acid cycle; Iron-sulfur;  
KW Mitochondrion; Transist peptide; 4fe-4s.  
FT TRANSIT 1  
MITOCHONDRION (POTENTIAL).

FT CHAIN 7 788 PROBABLE ACONITATE HYDRATASE.  
FT METAL 393 393 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 456 456 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 459 459 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 788 AA; 85712 MW; 8861E6FC198B70D9 CRC64;

Query Match 62.0%; Score 31; DB 1; Length 788;  
Best Local Similarity 66.7%; Pred. No. 1,4e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 SODVTAVA 11  
11:11 111  
Db 500 SPDITAVA 508

RESULT 14  
EGT2\_YEAST STANDARD; PRT; 1041 AA.  
ID EGT2\_YEAST  
AC P42835;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE EGT2 protein precursor (Early G1 transcript 2).  
GN EGT2 OR YNL327W OR N0320.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxId=4932;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1676;  
RX MEDLINE=95373280; PubMed=7645347;  
RA Martini M., Nicoud J.-M., Levesque H., Galliard C.;  
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV  
RT identifies the RPD3, PAS8 and KRL1 loci, five new open reading  
RT frames.";  
RL Yeast 11:567-572(1995).  
RN [2]  
RX SEQUENCE FROM N.A., AND FUNCTION.  
RX MEDLINE=96251274; PubMed=8668141;  
RA Kovacech B., Nasmyth K., Schuster T.;  
RT "EGT2 gene transcription is induced predominantly by Swi5 in early  
RT G1.";  
RL Mol. Cell. Biol. 16:3264-3274(1996).  
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL  
CC SEPARATION AFTER CYTOKINESIS. AS SEPARATION OF MUTANT DAUGHTER  
CC CELLS IS DELAYED, COULD EITHER BE AN ENZYME NECESSARY FOR GLUCANS-  
CC DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND  
CC DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC  
CC STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY  
CC G1; INACTIVATED BEFORE CELLS PASS START.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z46259; CAAB6371.1; -  
DR EMBL: Z71603; CAAG6259.1; -  
DR SGD: S0005271; EGT2.  
KW Glycoprotein; Repeat; Signal; Cell cycle; Cell division; Mitosis.  
FT SIGNAL 1 20  
FT CHAIN 21 1041  
FT CHAIN 200 203  
FT DOMAIN 381 384  
FT DOMAIN 388 395  
FT DOMAIN 490 493  
FT DOMAIN 586 589  
FT REPEAT 457 492  
EGT2 PROTEIN.  
POLY-SER.  
POLY-SER.  
POLY-SER.  
POLY-SER.  
POLY-THR.  
1-1.

```
FT REPEAT 577 606 1-2.
FT REPEAT 613 647 1-3.
FT REPEAT 716 745 1-4.
FT REPEAT 773 802 1-5.
FT REPEAT 811 840 1-6.
FT REPEAT 849 886 1-7.
FT REPEAT 887 924 1-8.
FT REPEAT 925 962 1-9.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1041 AA; 108494 MW; 01FECEFEAA8744CD CRC64;
```

Query Match 62.0%; Score 31; DB 1; Length 1041;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KASQDVT 8  
1:1:1:1:1  
DB 111 KSSDIT 118

```
RESULT 15
YA09_SCHPO
ID YA09_SCHPO STANDARD; PRT; 1275 AA.
AC 010164;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 143.6 kDa protein C26A3.09C in chromosome I.
GN SPAC26A3.09C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z69240; CA93232.1; -
CC HSSP; Q07960; IRGP.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000198; RHOGAP.
CC Pfam; PF00169; PH; 1.
```

```
DR Pfam; PF00620; RhogAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00324; RhogAP; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
KW Hypothetical protein
FT DOMAIN 719 836 PH.
SQ SEQUENCE 1275 AA; 143586 MW; 136DA1A26B4A1BBE CRC64;
```

Query Match 62.0%; Score 31; DB 1; Length 1275;  
Best Local Similarity 63.6%; Pred. No. 2.3e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KASQDVTAVA 11  
1:1:1:1:1  
DB 157 KSSKDVPTASA 167

Search completed: August 20, 2002, 11:41:20  
Job time: 399 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:41 ; Search time 43.57 Seconds

(without alignments)  
43.676 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_24\_34

Perfect score: 50

Sequence: 1 KASQDVTAAV 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues 562222

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-virus:\*
- 13: sp-vertebrate:\*
- 14: sp-unclassified:\*
- 15: sp-virus:\*
- 16: sp\_bacteriap:\*
- 17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	76.0	812	2	09RBA5
2	36	72.0	1449	2	068542
3	36	72.0	1449	2	055264
4	35	70.0	420	16	098NB5
5	34	68.0	245	2	09EUF7
6	34	68.0	264	10	09ATD1
7	34	68.0	328	16	0927J6
8	34	68.0	389	2	091604
9	34	68.0	395	5	017839
10	34	68.0	435	16	09A6D3
11	34	68.0	492	2	09AHT7
12	34	68.0	501	2	0924J8
13	34	68.0	553	5	09W458
14	34	68.0	584	2	09KH13
15	34	68.0	815	10	080623
16	34	68.0	1041	12	09E1F7

## ALIGNMENTS

17	34	68.0	1041	12	09DPQ2	09dpq2 meleagrid h
18	34	68.0	2146	3	059897	059897 aspergillus
19	34	68.0	2146	3	060026	060026 aspergillus
20	33	66.0	226	5	P91481	P91481 caenorhabdi
21	33	66.0	400	4	09NXH5	09nxh5 homo sapien
22	33	66.0	449	4	09H2K4	09h2k4 homo sapien
23	33	66.0	449	4	09F055	09f055 homo sapien
24	33	66.0	485	12	039258	039258 equine herp
25	33	66.0	501	4	0960U6	0960u6 homo sapien
26	33	66.0	501	4	0961X5	0961x5 homo sapien
27	33	66.0	721	4	09Y608	09y608 homo sapien
28	33	66.0	822	5	P91440	P91440 caenorhabdi
29	33	66.0	975	5	09XZU1	09xzul drosophila
30	33	66.0	975	5	09VJH4	09vjh4 drosophila
31	33	66.0	975	5	09UB14	09ub14 drosophila
32	33	66.0	2396	5	077291	077291 drosophila
33	32	64.0	117	10	091GCS	091gcs oryza sativ
34	32	64.0	138	11	09CPT6	09cpt6 mus musculu
35	32	64.0	181	16	098NFI	098nfi rhizobium l
36	32	64.0	252	10	09FJZ5	09fjz5 arabidopsis
37	32	64.0	263	16	09HVT4	09hvt4 pseudomonas
38	32	64.0	330	12	069497	069497 human herpe
39	32	64.0	348	12	066597	066597 epizootic h
40	32	64.0	355	5	09VTD4	09vtd4 drosophila
41	32	64.0	355	5	09NFT2	09nft2 drosophila
42	32	64.0	371	10	004342	004342 leishmania
43	32	64.0	385	5	09NLG2	09nlg2 caenorhabdi
44	32	64.0	398	5	027408	027408 caenorhabdi
45	32	64.0	482	1	P71412	P71412 halobacteri

RESULT 1

Q9RBA5 PRELIMINARY; PRT; 812 AA.

AC Q9RBA5; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE ENDO-INULINASE PRECURSOR (EC 3.2.1.7).

GN INUL1.

OS Archaeobacter sp. S37.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococccineae; Micrococccaceae; Archaeobacter.

OX NCBI\_TaxID=85703;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S37;

RA Kang S.I., Kim S.I.;

RT "Molecular cloning and sequence analysis of an endo-inulinase gene from Archaeobacter sp. S37."

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

CC -|- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

DR EMBL: AJ131562; CAB63119.1; -

DR InterPro: IPR001362; Glyco\_hydro\_32.

DR Pfam: PF00251; Glyco\_hydro\_32.

KW Glycosidase; Hydrolase; Signal.

FT SIGNAL 1

FT CHAIN 54 802

SQ SEQUENCE 812 AA; 87854 MW; 55A3E688A0C6B528 CRC64;

Query Match 76.0%; Score 38; DB 2; Length 812;

Best Local Similarity 80.0%; Pred. No. 42;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KASQDVTAAV 10

Db 243 KASQDVTAAV 252

```
RESULT 2
ID 068542 PRELIMINARY; PRT; 1449 AA.
AC 068542;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GLUCOSYLTRANSFERASE N (FRAGMENT).
GN GFN.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1477;
RA Jaffe R.I.;
RT "Streptococcus salivarius V1477 gtfN."
RL Submitted (FE8-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049609; AAC05156.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KM Transferase.
FT NON_TER
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;
```

```
Query Match 72.0%; Score 36; DB 2; Length 1449;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 2 ASODVTTAVA 11
Db 69 AAODPTTAVA 78
```

```
RESULT 3
ID 055264 PRELIMINARY; PRT; 1449 AA.
AC 055264;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUCOSYLTRANSFERASE PRECURSOR.
GN GTF.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=951212197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases."
RL Infect. Immun. 63:609-621(1995).
DR EMBL; L35495; AAC41412.1; -.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KM Signal; Transferase.
FT SIGNAL
KW CHAIN
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;
```

```
Query Match 72.0%; Score 36; DB 2; Length 1449;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 2 ASODVTTAVA 11
Db 69 AAODPTTAVA 78
```

```
RESULT 4
ID 098NB5 PRELIMINARY; PRT; 420 AA.
AC 098NB5;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE N-ACETYLIMORAMONYL-L-ALANINE AMIDASE.
GN MLR0213.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida T., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002994; BAB47846.1; -.
DR InterPro: IPR002508; Amidase_3.
DR Pfam: PF01520; Amidase_3; 1.
KM Complete proteome.
SQ SEQUENCE 420 AA; 44793 MW; 9FAFA1B6A998B89 CRC64;
```

```
Query Match 70.0%; Score 35; DB 16; Length 420;
Best Local Similarity 63.6%; Pred. No. 86;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 KASODVTTAVA 11
Db 397 KAAOSTINAVA 407
```

```
RESULT 5
ID 09EUF7 PRELIMINARY; PRT; 245 AA.
AC 09EUF7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PILX5 PROTEIN.
GN PILX5.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC Thesis (1998), Universidad de Cantabria, Departamento de Biología.
RA Nunez B.;
RL EMBL; AJ006342; CAC20142.1; -.
KM Plasmid.
SQ SEQUENCE 245 AA; 27705 MW; FB253DC89C368155 CRC64;
```

```
Query Match 68.0%; Score 34; DB 2; Length 245;
Best Local Similarity 54.5%; Pred. No. 78;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
```

OY 1 KASQDVTAV 11  
 Db 174 KATODATFNALS 184

## RESULT 6

Q9ATD1 PRELIMINARY; PRT; 264 AA.  
 AC Q9ATD1: 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 GN GHMYB9.  
 OS Gossypium hirsutum (Upland cotton).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Malvales; Malvaceae; Gossypium.  
 OX NCBL\_TaxID=3635;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ACALA MAXX; TISSUE=DAY OF ANTHESIS OVULE;  
 RA Matz E.C., Burr B.;  
 RT "Cotton seed fibers are trichomes."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF336286; AKI9619.1; -  
 DR HSSP; P01103; 1PCM.  
 DR InterPro; IPR001005; Myb\_DNA\_bind.  
 DR Pfam; PF00249; Myb\_DNA-binding; 2.  
 DR SMART; SM00395; SANT; 2.  
 DR PROSITE; PS00037; MYB\_1; 1.  
 DR PROSITE; PS00334; MYB\_2; 1.  
 DR PROSITE; PS50090; MYB\_3; 2.  
 SQ SEQUENCE 264 AA; 29629 MW; E09HCBB4E7662E95 CRC64;

Query Match 68.0%; Score 34; DB 10; Length 264;  
 Best Local Similarity 87.5%; Pred. No. 84;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KASQDVT 8  
 Db 131 EASQDVT 138

RESULT 7  
 Q927J6 PRELIMINARY; PRT; 328 AA.  
 AC Q927J6: 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE LIN2792 PROTEIN.  
 GN LIN2792.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Listeria.  
 OX NCBL\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RA Glaeser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Chabdit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussutget O.,  
 RA Eutlian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkhat G.,  
 RA Madeno E., Maitournam A., Mata Vicente J., Ng E., Negjari H.,  
 RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;

RT "Comparative genomics of Listeria species."  
 RL Science 294:849-852(2001).  
 DR EMBL; AL596173; CAC98018.1; -  
 DR ListLib; LIN02792; -  
 KW Complete proteome.  
 SQ SEQUENCE 328 AA; 36670 MW; F2E1A0F2C6D98622 CRC64;

Query Match 68.0%; Score 34; DB 16; Length 328;  
 Best Local Similarity 87.5%; Pred. No. 11e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 SODVTAV 10  
 Db 153 SEDVTAV 160

RESULT 8  
 Q9L6O4 PRELIMINARY; PRT; 389 AA.  
 AC Q9L6O4: 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE 89% IDENTITY WITH E. COLI PUTATIVE UROPOREHYRINOGEN III METHYLASE  
 DE (HEMX) (SP:p09127).  
 GN HEHX.  
 OS Salmonella typhimurium LT2.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBL\_TaxID=99287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SGSC1412;  
 RA Wastu;  
 RT "The Salmonella typhimurium Genome Sequencing Project."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SGSC1412;  
 RA Waterston R.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF233324; AAF33455.1; -  
 KW Methyltransferase.  
 SQ SEQUENCE 389 AA; 42285 MW; 25B73FA79157770E CRC64;

Query Match 68.0%; Score 34; DB 2; Length 389;  
 Best Local Similarity 70.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 ASQDVTAV 11  
 Db 151 SDQDVTAAA 160

RESULT 9  
 O17859 PRELIMINARY; PRT; 395 AA.  
 AC O17859: 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE F32H2.7 PROTEIN.  
 GN F32H2.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBL\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP Kershaw J.K.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "genome sequence of the nematode C.elegans: A platform for  
investigating biology";  
RL Science 282:2012-2018(1998).  
DR EMBL: Z81523; CAB04247.1; -.  
SQ SEQUENCE 395 AA; 43071 MW; 814EB82A7BE1160D CRC64;

Query Match 68.0%; Score 34; DB 5; Length 395;  
Best Local Similarity 70.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KASODVTTAV 10  
Db 365 RSSODVLTAV 374

RESULT 10  
Q9A6D3 PRELIMINARY; PRT; 435 AA.  
AC Q9A6D3;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DE D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
GN CC2161.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=69394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Niernan W.C., Feldlyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
Ratlova I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,  
Kolonyak J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL: AEO05888; AAK24132.1; -.  
DR MEROPS: S11; UPW: -.  
DR TIGR: CC2161; -.  
DR InterPro: IPR001967; DadacBptasel.  
DR Pfam: PF00768; Peptidase\_S11; 1.  
DR PRINTS: PR00725; DADACBPTASE1.  
KW Carboxypeptidase; Complete proteome.  
SQ SEQUENCE 435 AA; 46511 MW; 8A0DBD1971495B3E CRC64;

Query Match 68.0%; Score 34; DB 16; Length 435;  
Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KASODVTTAV 11  
Db 132 KSANDVATAVA 142

RESULT 11  
Q9AHT7 PRELIMINARY; PRT; 492 AA.  
AC Q9AHT7;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE 1,4-BETA-N-ACETYLMURAMIDASE (FRAGMENT).

GN LYTC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Witzmann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langemann S., Johnson S., Koenig S.;  
RT "Use of a Whole Genome Approach to Identify Vaccine Molecules  
Affording Protection against Streptococcus pneumoniae Infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL: AF291697; AAK19157.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 492 AA; 57554 MW; DF6DC1F1DA33B043 CRC64;

Query Match 68.0%; Score 34; DB 2; Length 492;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KASODVTTA 9  
Db 31 KTSODVTTA 39

RESULT 12  
Q9Z4J8 PRELIMINARY; PRT; 501 AA.  
AC Q9Z4J8;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE 1,4-BETA-N-ACETYLMURAMIDASE PRECURSOR (EC 3.2.1.17).  
GN LYTC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M31;  
RX MEDLINE=99340545; PubMed=10411730;  
RA Garcia P., Gonzalez M.P., Garcia E., Garcia J.L., Lopez R.;  
RT "The molecular characterization of the first autolytic lysozyme of  
Streptococcus pneumoniae reveals evolutionary mobile domains";  
RL Mol. Microbiol. 33:128-138(1999).  
DR EMBL: AJ009639; CAA08765.1; -.  
KW Signal; Hydrolase; Glycosidase.  
FT SIGNAL 1 33  
FT CHAIN 34 501  
SQ SEQUENCE 501 AA; 56682 MW; A6774B6BDF2EB704 CRC64;

Query Match 68.0%; Score 34; DB 2; Length 501;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KASODVTTA 9  
Db 40 KTSODVTTA 48

RESULT 13  
Q9W458 PRELIMINARY; PRT; 553 AA.  
AC Q9W458;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)  
 DE CG12236 PROTEIN.  
 GN CG12236.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aspayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslter C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jieqyam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mitsuhashi N.V., Mobarry C., Morris J., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AEO03435; AAF46100.1; -;  
 DR FlyBase: FBgn0029822; CG12236.  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00651; BTB; 1.  
 DR Pfam: PF00096; zif-C2H2; 2.  
 DR SMART: SM00225; BTB; 1.  
 DR SMART: SM00355; Znf-C2H2; 2.  
 DR PROSITE: PS50097; BTB; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 2.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 NCBI\_Taxid=7227; EB3019D282B19092 CRC64;  
 SO SEQUENCE 553 AA; 60801 MW; EB3019D282B19092 CRC64;

Query Match 68.0%; Score 34; DB 5; Length 553;  
 Best Local Similarity 63.6%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 KASODVTTAVA 11  
 DB 346 QTSODATTAA 356

RESULT 14  
 ID 09KH13 PRELIMINARY; PRT; 584 AA.  
 AC 09KH13;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)  
 DE ADHESIN PMGAL.4 (FRAGMENT).  
 GN PMGAL.4.  
 OS Mycoplasma gallisepticum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmataceae; Mycoplasma.  
 NCBI\_Taxid=2096;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HS;  
 RA Shen Q.C., Bi D.R., Weng C.J.;  
 RT "Sequence analysis of the pmga multigene family of Mycoplasma  
 RT gallisepticum strain HS.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF275312; AAF91415.1; -;  
 DR InterPro: IPR001986; BPSF\_synthase.  
 DR PROSITE: PS00104; BPSF\_SYNTHASE\_1; UNKNOWN\_1.  
 FT NON\_TER 584  
 FT SEQUENCE 584 AA; 62453 MW; 5C467BA55FB27A72 CRC64;  
 SO SEQUENCE 584 AA; 62453 MW; 5C467BA55FB27A72 CRC64;

Query Match 68.0%; Score 34; DB 2; Length 584;  
 Best Local Similarity 54.5%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 KASODVTTAVA 11  
 DB 224 KANEDIMTAVS 234

RESULT 15  
 ID 080623 PRELIMINARY; PRT; 815 AA.  
 AC 080623;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, last annotation update)  
 DE PUTATIVE SER/THR PROTEIN KINASE (HYPOTHETICAL 91.3 KDA PROTEIN).  
 GN F12I6.2 OR ATG39360.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_Taxid=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Symes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,  
 RA Somerville C.R., Venter J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F12I6 genomic sequence.";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Theologis A., Ecker J., Davis R.W.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AC004218; AAC27827.1; -;  
 DR EMBL: AF325084; AAK17152.1; -;  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR004040; STY\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR SMART: SM00221; STYKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;  
KW Transferase.  
SQ SEQUENCE 815 AA; 91321 MW; D3F4E07B73D8342D CRC64;

Query Match 68.0%; Score 34; DB 10; Length 815;  
Best Local Similarity 87.5%; Pred. NO. 2.7e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ASQDVTTA 9  
|||:|  
Db 110 ASQDUTTA 117

Search completed: August 20, 2002, 11:40:53  
Job time: 432 sec



PR 10-MAY-1996; 96US-0017466.  
XX (BIOJ ) BIOGEN INC.  
PA  
XX Benjamin CD, Burkly LC, Hession C, Whitty A;  
PI  
XX WPI: 1998-008885/01.  
DR N-PSDB: AAT97440.  
XX  
XX Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
XX Claim 22: Page 81; 111pp; English.  
PS  
XX This polypeptide comprises the light chain variable region (VL) of  
CC monoclonal antibody (Mab) C9.B8, which is produced by a hybridoma  
CC deposited as ATCC 12107, and which is specific for the gamma  
CC constant (gc) chain (see AAW31646) of human cytokine receptors. The  
CC invention provides compositions and methods for inhibiting cytokine  
CC signalling using gc chain blocking agents for the treatment of  
CC immunological diseases such as myasthenia gravis, rheumatoid  
CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
CC allergy, asthma, parasitic infection, graft vs. host disease or  
CC psoriasis. Preferred gc blocking agents include Mab CP.B8, its Fab  
CC fragment and an antibody having a light chain variable region  
CC CDR selected from those of CP.B8 VL or a heavy chain variable  
CC region CDR selected from those of CP.B8 VH (see AAW31648).  
XX  
SQ Sequence 106 AA;

Query Match 100.0%; Score 42; DB 19; Length 106;  
Best Local Similarity 100.0%; Pred. NO. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WASTRHT 7  
| | | | | | |  
DB 50 wastrht 56

RESULT 2  
AAR25723  
ID AAR25723 standard; Protein: 107 AA.  
XX  
AC AAR25723;  
XX

DT 13-JAN-1993 (first entry)  
XX

DE Humanised VL region of the mouse Fd 138-80 antibody.  
XX

XX Murine: immunoglobulin; CDR; HSV; non immunogenic; herpes simplex  
KW virus; HSV; light chain; variable region; framework; human; Eu.  
XX

OS Mus musculus.  
XX

XX Key Location/Qualifiers  
FH Region 24..38  
FT /note= "CDR"

FT Region 50..56  
FT /note= "CDR"

FT Region 89..97  
FT /note= "CDR"

FT Misc-difference 36  
FT /note= "mutated residue"

FT Misc-difference 48  
FT /note= "mutated residue"

FT Misc-difference 63  
FT /note= "mutated residue"

FT Misc-difference 87  
FT /note= "mutated residue"

XX

PN W09211018-A.  
XX  
XX 09-JUL-1992.  
PD  
XX 19-DEC-1991; 91WO-US09711.  
XX  
XX 19-DEC-1990; 90US-0634278.  
PR  
XX (PROT-) PROTEIN DESIGN LABS INC.  
PA  
XX Co MS, Coellnigh KL, Landolf NF, Queen CL, Schneider WP;  
PI  
XX WPI: 1992-249842/30.  
DR  
XX  
XX New immunoglobulin(s) having murine CDRs in human framework  
PT regions - have lower antigenicity; useful for treating e.g. HSV,  
PT CMV, T-cell disorders, myeloid disorders and auto-immune  
PT conditions  
XX  
XX Claim 19: Fig 14D; 141pp; English.  
PS  
XX The sequence shows the humanised mature light chain variable region  
CC of the mouse Fd 138-80 antibody. Murine CDRs were used in a  
CC human Eu framework to produce a pure humanised immunoglobulin (Ig)  
CC which is capable of binding to a herpes simplex virus-specific  
CC epitope. The Ig is non immunogenic, due to the human framework,  
CC and has a strong affinity for its predetermined antigen. They can  
CC be produced in large quantities via recombinant DNA and monoclonal  
CC antibody technology. The humanised Igs may be used alone or in  
CC combination with chemotherapeutic agents such as non-steroidal  
CC anti-inflammatory drugs or immunosuppressants.  
XX  
SQ Sequence 107 AA;

Query Match 100.0%; Score 42; DB 13; Length 107;  
Best Local Similarity 100.0%; Pred. NO. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WASTRHT 7  
| | | | | | |  
DB 50 wastrht 56

RESULT 3  
AAR55121  
ID AAR55121 standard; Protein: 107 AA.  
XX  
AC AAR55121;  
XX

DT 27-JAN-1995 (first entry)  
XX

DE Mouse anti-HIV mu39.1 light chain variable region.  
XX

XX Immunoglobulin; heavy chain; anti-HIV antibody; neutralisation;  
KW human immunodeficiency virus; variable region; VH chain; murine.  
XX

OS Mus musculus.  
XX

XX Key Location/Qualifiers  
FH Region 1..23  
FT /label= FR1

FT Region 24..34  
FT /label= CDRL1

FT Region 35..49  
FT /label= FR2

FT Region 50..56  
FT /label= CDR2

FT Region 57..88  
FT /label= FR3

FT Region 89..97  
FT /label= CDR3



```

FT Region 98..107.
FT /label= FR4
XX
XX JP06125783-A.
XX
XX 10-MAY-1994.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX (KAGA-) ZH KAGAKU KESSEI-RYOHO KENKYUSHO.
XX
XX WPI; 1994-187942/23.
XX
XX N-PSDB; AAO65552.
XX
XX Mouse-human chimeric anti-HIV antibody heavy and light chains -
XX and recombinant antibody consisting of the H- and L-chains,
XX useful in AIDS therapy
XX
XX Example 3; Fig 2; 22pp; Japanese.
XX
XX Murine monoclonal antibodies mu 39.1 and mu 5.5 both neutralise HIV.
XX The heavy and light chain variable regions from these antibodies
XX were sequenced (AAR55120-R55123). The murine anti-HIV CDRs were
XX introduced into human framework regions to construct chimeric
XX antibodies (AAR55124-R55127).
XX
XX Sequence 107 AA:
SQ

```

```

Query Match 100.0%; Score 42; DB 15; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 WASTRHT 7
   |||||
DB 50 wastrht 56

```

```

RESULT 4
AAR55125 ID AAR55125 standard; Protein; 107 AA.
XX
XX AAR55125;
AC
XX
DT 30-JAN-1995 (first entry)
XX
DE Mouse-human chimeric anti-HIV mu39.1-derived light chain V region.
XX
XX Immunoglobulin; heavy chain; anti-HIV antibody; neutralisation;
XX human immunodeficiency virus; variable region; VL chain; murine;
XX chimeric; humanised.
OS Chimeric Mus musculus.
OS Chimeric Homo sapiens.
XX
XX Key Location/Qualifiers
FT 1..23 /label= FR1
FT /note= "human"
FT Region 24..34 /label= CDR1
FT /note= "murine"
FT Region 35..49 /label= FR2
FT /note= "human, except for murine residue Ser49"
FT Region 50..56 /label= CDR2
FT /note= "murine"
FT Region 57..88 /label= FR3
FT /note= "human"
FT

```

```

FT Region 89..97
FT /label= CDR3
FT /note= "murine"
XX
XX Region 98..107 /label= FR4
FT /note= "human"
XX
XX JP06125783-A.
XX
XX 10-MAY-1994.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX 28-DEC-1991; 91JP-0359808.
XX
XX (KAGA-) ZH KAGAKU KESSEI-RYOHO KENKYUSHO.
XX
XX WPI; 1994-187942/23.
XX
XX N-PSDB; AAO65556.
XX
XX Mouse-human chimeric anti-HIV antibody heavy and light chains -
XX and recombinant antibody consisting of the H- and L-chains,
XX useful in AIDS therapy
XX
XX Claim 5; Fig 10; 22pp; Japanese.
XX
XX Murine monoclonal antibodies mu 39.1 and mu 5.5 both neutralise HIV.
XX The heavy and light chain variable regions from these antibodies
XX were sequenced (AAR55120-R55123). The murine anti-HIV CDRs were
XX introduced into human framework regions to construct chimeric
XX antibodies (AAR55124-R55127).
XX
XX Sequence 107 AA:
SQ

```

```

Query Match 100.0%; Score 42; DB 15; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 WASTRHT 7
   |||||
DB 50 wastrht 56

```

```

RESULT 5
AAR60300 ID AAR60300 standard; Protein; 107 AA.
XX
XX AAR60300;
AC
XX
DT 09-MAR-1995 (first entry)
XX
DE Anti HIV antibody light chain variable region.
XX
XX Antibody; heavy chain; light chain; human immunodeficiency virus;
XX HIV; acquired immune deficiency syndrome; AIDS; treatment;
XX prophylaxis; Mus musculus; Homo sapiens.
XX
XX Mus musculus.
OS
XX
XX Key Location/Qualifiers
FT 1..23 /label= Framework region 1.
FT Region 24..34 /label= CDR1.
FT Region 35..49 /label= Framework region 2.
FT Region 50..56 /label= CDR2.
FT Region 57..88 /label= Framework region 3.
FT Region 89..97 /label= CDR3.
FT

```

FT	Region	98..107
FT	/label= Framework region 4.	
XX		
PN	W09415969-A.	
XX		
PD	21-JUL-1994.	
XX		
PE	14-JAN-1993; 93WO-JP00039.	
XX		
PR	14-JAN-1993; 93AU-0032671.	
PR	14-JAN-1993; 93WO-JP00039.	
PA	(KAGA ) CHERO SERO THERAPEUTIC RES INST.	
XX		
PI	Eda Y, Kimachi K, Maeda H, Osatomi K, Shiosaki K;	
PI	TOKiyoshi S;	
XX		
DR	WPI: 1994-249145/30.	
DR	N-PSDB; AAQ70370.	
XX		
PT	Recombinant chimeric anti HIV antibody - useful for the treatment and prevention of HIV	
XX		
PS	Claim 7; Figure 2; 51pp; Japanese.	
CC	The recombinant antibody light chain has neutralising activity against HIV. Chimeric antibodies comprising both mouse and human sequences are useful in the treatment/prevention of AIDS caused by HIV.	
CC		
SO	Sequence 107 AA;	
OY	Query Match 100.0%; Score 42; DB 15; Length 107; Best Local Similarity 100.0%; Pred. No. 0.72; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0. 1 WASTRHT 7         50 wastrht 56	
Db		
RESULT 6		
AAR60304		
ID AAR60304 standard; Protein: 107 AA.		
XX		
AC AAR60304;		
XX		
DT 13-MAR-1995 (first entry)		
DE Chimeric anti HIV antibody light chain variable region.		
XX		
KW Antibody; heavy chain; light chain; human immunodeficiency virus; HIV; acquired immune deficiency syndrome; AIDS; treatment; prophylaxis; Mus musculus; Homo sapiens.		
KW		
XX		
OS Chimeric Homo sapiens.		
OS Chimeric Mus musculus.		
XX		
Key Location/Qualifiers		
FT Region 1..23		
FT /label= Framework region 1.		
FT Region 24..34		
FT /label= CDRL.		
FT /note= "Mouse derived amino acid sequence."		
FT Region 35..49		
FT /label= Framework region 2.		
FT Region 49..56		
FT /note= "Mouse derived amino acid sequence."		
FT Region 50..56		
FT /label= CDR2.		
FT Region 57..88		
FT /label= Framework region 3.		

```

FT      Region          89..97  

FT      /label= CDR3.  

FT      /note= "Mouse derived amino acid sequence."  

FT      Region          98..107  

FT      /label= Framework region 4.  

XX      WO9415969-A.  

XX      21-JUL-1994.  

XX      PD      14-JAN-1993; 93WO-JP00039.  

XX      PE      14-JAN-1993; 93AU-0032671.  

XX      PR      14-JAN-1993; 93WO-JP00039.  

XX      (KAGA ) CHEMO SERO THERAPEUTIC RES INST.  

XX      XX      Eda Y, Kimachi K, Maeda H, Osatomi K, Shiosaki K;  

XX      PI      Tokiyoshi S;  

XX      WPI: 1994-249145/30.  

XX      DR      N-PSDB; AAQ70374.  

XX      PT      Recombinant chimeric anti HIV antibody - useful for the treatment  

XX      PS      and prevention of HIV  

XX      Claim 6; Figure 10; 51pp; Japanese.  

XX      CC      The recombinant antibody light chain has neutralising activity  

XX      CC      against HIV. Chimeric antibodies comprising both mouse and human  

XX      CC      sequences are useful in the treatment/prevention of AIDS caused by  

XX      CC      HIV.  

XX      SO      Sequence 107 AA;  

  

Query Match          100.0%; Score 42; DB 15; Length 107;  

Best Local Similarity 100.0%; Pred. NO. 0.72;  

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  

QY      1 WASTRHT 7  

        |||||  

Db      50 wastrht 56  

  

RESULT 7  

AAW47087  

ID      AAW47087 standard; Protein; 107 AA.  

XX      AC      AAW47087;  

XX      DT      26-JUN-1998 (first entry)  

XX      DE      Mouse J591 monoclonal antibody light chain variable region VK17.  

XX      KW      Mouse; monoclonal antibody; J591; prostate specific membrane antigen;  

XX      KW      cancer; vascular endothelial cell; metastatic adenocarcinoma.  

XX      OS      Mus sp.  

XX      WO9803873-A1.  

XX      PD      29-JAN-1998.  

XX      PE      17-JUL-1997; 97WO-US12035.  

XX      PR      09-APR-1997; 97US-0838682.  

XX      PR      18-JUL-1996; 96US-0022125.  

XX      PA      (CORR ) CORNELL RES FOUND INC.  

XX      PI      Bander NH;  

XX

```

DR WPI: 1998-120937/11.  
DR N-PSDB; AAV13953.

XX Destroying cancer cells with agent that binds to prostate specific  
PT membrane antigen - on vascular endothelial cells near the cancer, or  
PT on normal, hypertrophic or cancerous prostatic cells, also used for  
PT diagnosis

PS Example 12: Page 61; 94pp; English.

XX The present sequence represents the mouse J591 monoclonal antibody light  
CC chain variable region from clone VK17 from an example of the present  
CC invention. The present invention describes the elimination of cancer  
CC cells by treating vascular endothelial cells (VEC) close to the cancer  
CC with an agent (A) able to bind to the extracellular domain (ECD) of  
CC prostate specific membrane antigen (PSMA). (A) both binds to the VEC and  
CC destroys the cancer cells. Also described are: (1) the detection of  
CC cancer tissue by detecting binding of labelled (A) to VEC close to, or  
CC within, a cancer tissue; (2) eliminating or detecting normal, benignly  
CC hyperplastic or cancerous prostate epithelial cells using optionally  
CC labelled (A); (3) hydridomas that produce a monoclonal antibody (Mab)  
CC that binds to PSMA. The method is used to treat renal, urothelial,  
CC colon, lung, rectal or breast cancers and metastatic adenocarcinoma of  
CC the liver. The diagnostic method is particularly used to detect  
CC recurrence of prostatic disease or to monitor the effect of treatments  
CC for prostate cancer (presence of PSMA in the serum indicates that  
CC prostate cells are being lysed). (A) binds to an epitope of PSMA  
CC expressed on live cells (contrast antibody 7B11 which only binds after  
CC cell lysis), allowing targeting of live, unfixed cells and thus  
CC providing more efficient treatment and diagnosis. Both cancer cells  
CC themselves and the VEC on which they depend are killed. All VEC close to  
CC cancer cells express PSMA, whatever the type of cancer, but normal VEC  
CC do not.

XX Sequence 107 AA;

Query Match 100.0%; Score 42; DB 19; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
| | | | |  
Db 50 wastrht 56

RESULT 8

ID AAY90375 standard; Protein; 107 AA.

XX AAY90375;

DT 15-JAN-2001 (first entry)

XX J591 monoclonal antibody light chain protein sequence fragment.

DE J591 monoclonal antibody: extracellular domain; diagnosis; therapy;

KW J591 monoclonal antibody: extracellular domain; diagnosis; therapy;

KM prostate specific membrane antigen; prostate cancer; light chain.

XX Homo sapiens.

OS US6107090-A.

XX US6107090-A.

XX 22-AUG-2000.

XX 09-APR-1997; 97US-0838682.

XX 06-MAY-1996; 96US-0016976.

XX 18-JUL-1996; 96US-0022125.

PA (CORR ) CORNELL RES FOUND INC.

XX Bander NH;

XX WPI: 2000-571325/53.  
DR N-PSDB; AAA37835.

XX Antibody specific for extracellular prostate-specific membrane antigen,  
PT useful for diagnosis and treatment of prostate cancer -  
PT

PS Example 12: Column 24; 33pp; English.

XX This sequence is a fragment of the light chain of the monoclonal antibody  
CC J591. The invention relates to an isolated antibody or its antigen  
CC binding portion (I) which binds to an extracellular domain of prostate  
CC specific membrane antigen and which does not require cell lysis to bind  
CC to the extracellular domain. The antibody or its antigen binding portion  
CC is selected for its ability to bind to live cells. (I) is useful for  
CC diagnosis of diseases associated with the presence of normal, benign  
CC hyperplastic, and cancerous epithelial cells or portions. Also it can be  
CC used for identifying the recurrence of such diseases, particularly when  
CC the disease is localised in a particular biological material of the  
CC patient for e.g. recurrence of prostatic disease. They can also be used  
CC alone or bound to a substance effective to kill cancerous prostate  
CC epithelial cells as a therapy for prostate cancer. Binding and  
CC internalising of the antibody with the prostate specific membrane  
CC antigen, permits the therapeutic use of intracellularly acting cytotoxic  
CC agents. (I) targets only prostate epithelial cells and other tissue are  
CC spared which provides safer treatment particularly for elderly patients.  
CC The antibodies bind to living prostate cells and treatments using these  
CC antibodies are more effective than those which target lysed prostate  
CC cells.

XX Sequence 107 AA;

Query Match 100.0%; Score 42; DB 21; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
| | | | |  
Db 50 wastrht 56

RESULT 9

ID AAB69665 standard; Protein; 107 AA.

XX AAB69665;

DT 30-APR-2001 (first entry)

XX Murine Fd138-80 antibody light chain SEQ ID NO: 50.

XX Humanised immunoglobulin; mouse; human; antibody; heavy chain; diabetes;

KW light chain; graft versus host disease; transplant; autoimmune disease;

KM multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;

XX myasthenia gravis; herpes infection; myeloid leukaemia; CMV infection.

XX Mus sp.

OS US6180370-B1.

XX US6180370-B1.

XX 30-JAN-2001.

XX 07-JUN-1995; 95US-0484537.

XX 28-DEC-1988; 88US-0290975.

XX 13-FEB-1989; 89US-0310252.

XX 28-SEP-1990; 90US-0590274.

XX 19-DEC-1990; 90US-0634278.

PA (PROT-) PROTEIN DESIGN LABS INC.

XX Queen CL, Selick HE;

XX WPI: 2001-190856/19.  
XX  
XX Producing humanized immunoglobulin, involves producing a cell  
PT containing DNA segments encoding humanized heavy and light chain  
PT variable regions, and expressing the DNA segments in the cell -  
XX  
XX Example 6; Fig 3; 145pp; English.  
XX  
XX The present invention describes a method of producing humanised  
CC immunoglobulins involving expressing in a cell a nucleic acid encoding a  
CC humanised version of an immunoglobulin. This is obtained by comparing a  
CC donor and human immunoglobulin and producing a combined antibody which  
CC contains part of each. These are useful in the treatment of  
CC graft-versus-host disease, transplant rejection, autoimmune diseases such  
CC as diabetes, rheumatoid arthritis, myasthenia gravis, multiple sclerosis  
CC and systemic lupus erythematosus, herpes infections, CMV virus infections  
CC and myeloid leukaemia. The present sequence is an antibody used to  
CC demonstrate the method of the invention.  
XX  
XX Sequence 107 AA;  
SQ

Query Match 100.0%; Score 42; DB 22; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WASTRHT 7  
| | | | | | |  
Db 50 wastrht 56

RESULT 10  
AAB69666  
ID AAB69666 standard; Protein; 107 AA.  
XX

AC AAB69666;  
XX

DT 30-APR-2001 (first entry)  
XX

DE Humanised Fd138-80 antibody light chain SEQ ID NO: 51.  
XX

XX Humanised immunoglobulin; mouse; human; antibody; heavy chain; diabetes;  
KK light chain; graft versus host disease; transplant; autoimmune disease;  
KM multiple sclerosis; rheumatoid arthritis; systemic lupus erythematosus;  
XX myasthenia gravis; herpes infection; myeloid leukaemia; CMV infection.  
XX

OS Mus sp.  
OS Homo sapiens.  
XX

PN US6180370-B1.  
XX

PD 30-JAN-2001.  
XX

PE 07-JUN-1995; 95US-0484537.  
XX

PR 28-DEC-1988; 88US-0290975.  
XX

PR 13-FEB-1989; 89US-0310252.  
XX

PR 28-SEP-1990; 90US-0590274.  
XX

PR 19-DEC-1990; 90US-0634278.  
XX

PA (PROT-) PROTEIN DESIGN LABS INC.  
XX

PI Queen CL, Selick HE;  
XX

DR WPI: 2001-190856/19.  
XX

XX Producing humanized immunoglobulin, involves producing a cell  
PT containing DNA segments encoding humanized heavy and light chain  
PT variable regions, and expressing the DNA segments in the cell -  
XX  
XX Example 6; Fig 3; 145pp; English.  
XX

CC The present invention describes a method of producing humanised  
CC immunoglobulins involving expressing in a cell a nucleic acid encoding a  
CC humanised version of an immunoglobulin. This is obtained by comparing a  
CC donor and human immunoglobulin and producing a combined antibody which  
CC contains part of each. These are useful in the treatment of  
CC graft-versus-host disease, transplant rejection, autoimmune diseases such  
CC as diabetes, rheumatoid arthritis, myasthenia gravis, multiple sclerosis  
CC and systemic lupus erythematosus, herpes infections, CMV virus infections  
CC and myeloid leukaemia. The present sequence is an antibody used to  
CC demonstrate the method of the invention.  
XX  
XX Sequence 107 AA;  
SQ

Query Match 100.0%; Score 42; DB 22; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WASTRHT 7  
| | | | | | |  
Db 50 wastrht 56

RESULT 11  
AAB36226  
ID AAB36226 standard; Protein; 107 AA.  
XX

AC AAB36226;  
XX

DT 15-FEB-2001 (first entry)  
XX

DE Monoclonal antibody J591 kappa light chain V region #2.  
XX

XX Mouse; antibody; heavy chain; prostate cancer; biological agent.  
XX

OS Mus sp.  
XX

PN US6136311-A.  
XX

PD 24-OCT-2000.  
XX

PE 17-JUL-1997; 97US-0895914.  
XX

PR 06-MAY-1996; 96US-0016976.  
XX

PR 18-JUL-1996; 96US-0022125.  
XX

PR 09-APR-1997; 97US-0838682.  
XX

PA (CORR ) CORNELL RES FOUND INC.  
XX

PI Bander NH;  
XX

DR WPI: 2001-040234/05.  
XX

DR N-PSDB; AAC66546.  
XX

XX Use of E99, a J415, a J533 or a J591 monoclonal antibodies for ablating  
PT or killing cancerous, especially non-prostate, cells (e.g. breast  
PT cancerous cells or cancerous cells of metastatic adenocarcinoma to the  
PT liver)  
XX

PS Example 12; Column 26; 35pp; English.  
XX

XX The present invention describes a method of killing cancer cells,  
CC particularly prostate cancer cells, by directing a biological agent to  
CC the cells which then binds to a prostate specific membrane antigen and  
CC causes the molecule to be internalised. The internalisation of the agent,  
CC which may be bound to a drug or which may act to kill the cell alone,  
CC then leads to the death of the cell. The present sequence forms  
CC part of an antibody which may be used as the biological agent of the  
CC invention. In addition to prostate cancer, the method can be used with  
CC renal, urothelial, colon, renal, lung and breast cancer cells, and  
CC cancerous cells of metastatic adenocarcinoma to the liver.  
XX

SQ Sequence 107 AA;  
XX

Query Match 100.0%; Score 42; DB 22; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.72; Mismatches 7; Conservative 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
 |||||  
 Db 50 wastrht 56

RESULT 12  
 AAY21817  
 ID AAY21817 standard; Protein; 108 AA.  
 XX  
 AC AAY21817;  
 XX  
 DT 10-SEP-1999 (first entry)  
 XX  
 DE Anti-STX1 light chain variable region.  
 XX  
 KM Humanised: monoclonal antibody; MAb: Shiga toxin; immunoglobulin;  
 KM Ig constant region; enterohemorrhagic Escherichia coli; EHEC; uremia;  
 KM edema; bloody diarrhoea; hemorrhagic colitis; hemolytic uremic syndrome;  
 KM thrombocytopenia; EHEC-mediated disease; anti-STX1.  
 XX  
 OS Shigella dysenteriae.  
 XX  
 FH Location/Qualifiers  
 FT Region 24..34  
 FT /note= "complementarity determining region (CDR) 1;  
 FT specifically claimed for in claim 10"  
 FT Region 50..56  
 FT /note= "complementarity determining region (CDR) 2;  
 FT specifically claimed for in claim 10"  
 FT Region 89..97  
 FT /note= "complementarity determining region (CDR) 3;  
 FT specifically claimed for in claim 10"  
 XX  
 PN MO9932645-A1.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PF 22-DEC-1998; 98WO-0527267.  
 XX  
 PR 18-DEC-1998; 98US-0215163.  
 PR 23-DEC-1997; 97US-0068635.  
 XX  
 PA (MELT/) MELTON-CELSA A.  
 PA (OBRI/) O'BRIEN A D.  
 PA (SCHM/) SCHMITT C K.  
 PA (STIN/) STINSON J L.  
 PA (WONG) WONG H.  
 XX  
 PI Melton-Celsa A, O'Brien AD, Schmitt CK, Stinson JL;  
 PI Wong H;  
 XX  
 DR WPI: 1999-418935/35.  
 DR N-PSDB: AAX82029.  
 XX  
 XX Humanized monoclonal antibodies against Shiga toxins, useful for  
 PT protection against enterohemorrhagic Escherichia coli or other Shiga  
 PT toxin producing bacteria  
 XX  
 PS Claim 5; Fig 3; 75pp; English.  
 XX  
 CC The invention relates to humanised monoclonal antibodies (MAb) against  
 CC Shiga toxins. The humanised MAb that binds to Shiga toxin comprises a  
 CC constant and a variable region, where: (a) the constant region contains  
 CC at least part of a human immunoglobulin (Ig) constant region; and (b) the  
 CC variable region contains at least part of a non-human Ig variable region.  
 CC Host cells transformed with vectors encoding a humanised MAb against  
 CC Shiga toxin type 2 is useful for treating a patient with an infection

CC caused by enterohemorrhagic Escherichia coli (EHEC) or other Shiga toxin  
 CC producing bacteria. The humanised MAb can also be used to reduce illness  
 CC caused by EHEC or other Shiga toxin producing bacteria. EHEC are  
 CC associated with food-borne outbreaks of bloody diarrhoea (hemorrhagic  
 CC colitis) and the hemolytic uremic syndrome. In particular, the humanised  
 CC MAb ameliorate edema, thrombocytopenia and uremia associated with EHEC-  
 CC mediated disease. The present sequence represents an anti-STX1 light  
 CC chain variable region.  
 XX  
 SQ Sequence 108 AA;

Query Match 100.0%; Score 42; DB 20; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 0.73; Mismatches 7; Conservative 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
 |||||  
 Db 50 wastrht 56

RESULT 13  
 AAP40031  
 ID AAP40031 standard; Protein; 223 AA.  
 XX  
 AC AAP40031;  
 XX  
 DT 01-DEC-1991 (first entry)  
 XX  
 DE Kappa anti-carcinoembryonic antigen chain.  
 XX  
 KM Immunoglobulin; carcinoembryonic antigen.  
 XX  
 PN EP125023-A.  
 XX  
 PD 14-NOV-1984.  
 XX  
 PF 06-APR-1984; 84EP-0302368.  
 XX  
 PR 08-APR-1983; 83US-0483457.  
 XX  
 PA (CITY ) CITY OF HOPE.  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Cabilly S, Holmes WE, Wetzel RB, Heyneker HL, Riggs AD;  
 XX  
 DR WPI: 1984-283749/46.  
 DR N-PSDB: AANA0023.  
 XX  
 PT Immunoglobulin(s) produced by recombinant host cells - useful as  
 PT antibodies analogous to forms from mammals.  
 XX  
 PS Disclosure; Fig. 3; 79pp; English.  
 XX  
 CC The protein is encoded by the mRNA coding sequence of the cDNA fragment  
 CC contained within recombinant vector pK17G4, using the vector this  
 CC immunoglobulin is produced readily in pure monoclonal form. Genetic  
 CC manipulations can be used to produce chimeras of variants drawing their  
 CC homology from species differing from each other. Protein manipulation  
 CC is also possible.  
 XX  
 SQ Sequence 223 AA;

Query Match 100.0%; Score 42; DB 5; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 1.5; Mismatches 7; Conservative 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
 |||||  
 Db 59 wastrht 65

RESULT 14  
AAP80155  
ID AAP80155 standard; protein; 247 AA.  
XX  
AC AAP80155;  
XX  
DT 13-OCT-1990 (first entry)  
XX  
DE FRS and CDRs from anti-CEA monoclonal antibody.  
XX  
KM Biosynthetic Antibody Binding site (BABS); variable domain;  
KW anti-CEA monoclonal antibody; ss.  
XX  
OS Synthetic.  
XX  
PN WO8809344-A.  
XX  
PD 01-DEC-1988.  
XX  
FE 19-MAY-1988; 88WO-US01737.  
XX  
PR 21-MAY-1987; 87US-0052800.  
XX  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
PI Huston JS, Oppermann H;  
XX  
DR WPI: 1988-353928/49.  
DR N-PSDB; AAN80181.  
XX  
PT Recombinant multifunctional protein - having an Ab binding site and a  
PT sequence for biological activity, ion sequestering or binding to a solid  
PT support.  
XX  
PS Disclosure; : 15pp; English.  
XX  
CC Multi-functional biosynthetic protein comprising FRS and  
CC CDRs with specificity for a tumour antigen CEA.  
CC See also AAN80171-N80192.  
XX  
SQ Sequence 247 AA;  
OY 1 WASTRHT 7  
DB 190 wastrht 196

Query Match 100.0%; Score 42; DB 9; Length 247;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 15  
AAP80156  
ID AAP80156 standard; protein; 247 AA.  
XX  
AC AAP80156;  
XX  
DT 14-NOV-1990 (first entry)  
XX  
DE Biosynthetic antibody binding site.  
XX  
KM Biosynthetic antibody binding site; framework region; assay; imaging;  
KW multifunctional protein.  
XX  
PN WO8809344-A.  
XX  
PD 01-DEC-1988.  
XX  
PF 19-MAY-1988; 88WO-US01737.  
XX  
PR 21-MAY-1987; 87US-0052800.  
XX

PA (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
PI Huston JS, Oppermann H;  
XX  
DR WPI: 1988-353928/49.  
DR N-PSDB; AAN80182.  
XX  
PT Recombinant multifunctional protein - having antibody binding site and a  
PT sequence for biological activity, ion sequestering or binding to a  
PT solid support.  
XX  
PS Disclosure; : 15pp; English.  
XX  
CC The biosynthetic antibody binding site forms part of a recombinant  
CC multifunctional protein which also comprises an effector protein, an AA  
CC sequence capable of sequestering an ion, or a sequence capable of  
CC binding to a solid support. The BABS comprises framework regions in both  
CC VH and LH and CDR1 and CDR3 in VH, and CDR1, CDR2 and CDR3 in VL from an  
CC anti-CEA monoclonal antibody. CDR2 in VH is a CDR2 consensus sequence  
CC found in most immunoglobulin VH regions. The effector protein is an  
CC enzyme, toxin, receptor, binding site, growth factor, cytokine  
CC or antimetabolite. The sequence capable of sequestering an ion is  
CC calmodulin or metallothionein. The sequence able to bind to a solid  
CC support is streptavidin or a protein A fragment. The protein may be used  
CC for, eg specific binding assays, affinity purification, biocatalysts, drug  
CC targeting, imaging and immunological treatment of oncogenic and  
CC infectious diseases. The protein offers fewer cleavage sites to  
CC circulating proteolytic enzymes and have improved stability.  
CC They reach target organs rapidly and are cleared quickly from the body.  
XX  
SQ Sequence 247 AA;  
OY 1 WASTRHT 7  
DB 190 wastrht 196

Query Match 100.0%; Score 42; DB 9; Length 247;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: August 20, 2002, 11:34:39  
Job time: 633 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:35:17 ; Search time 25.74 Seconds  
(without alignments)  
26.132 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_50\_56

Perfect score: 42  
Sequence: 1 WASTRHT 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	117	2 S42466	Ig kappa chain V r
2	42	100.0	131	2 P10207	anti-idiotypic ant
3	42	100.0	152	2 S30751	Ig kappa chain pre
4	34	81.0	212	2 T27834	hypothetical prote
5	34	81.0	246	2 C38113	yebC protein - Esc
6	34	81.0	246	2 F90950	hypothetical prote
7	34	81.0	246	2 B85799	hypothetical prote
8	34	81.0	246	2 AG0743	conserved hypotnet
9	34	81.0	605	2 T49780	related to beta tr
10	34	81.0	3010	1 GNMWTW	genome polypotein
11	33	78.6	206	2 C83590	probable transcript
12	33	78.6	212	2 C90786	probable tet opero
13	33	78.6	212	2 C64843	probable transcrip
14	33	78.6	212	2 A85646	probable tet opero
15	33	78.6	275	2 D97525	hypothetical prote
16	33	78.6	347	2 D83554	hypothetical prote
17	33	78.6	380	2 S62525	hypothetical prote
18	33	78.6	1139	2 S44800	FI0E9.8 protein -
19	32	76.2	142	2 C82829	conserved hypotnet
20	32	76.2	268	2 E86962	probable oxidoredu
21	32	76.2	268	2 E70551	probable oxidoredu
22	32	76.2	434	2 AG1697	hypothetical prote
23	32	76.2	572	2 DEECPC	pyruvate dehydroge
24	32	76.2	572	2 E90748	pyruvate oxidase I
25	32	76.2	572	2 A85599	pyruvate oxidase P
26	32	76.2	725	2 D73284	hypothetical prote
27	32	76.2	738	2 T01897	ethylene-response
28	32	76.2	745	2 E75354	DNA helicase II -
29	32	76.2	889	2 A41259	potassium transpor

30	32	76.2	1330	1 VCBES5	major capsid prote
31	32	76.2	1385	2 S61236	major capsid prote
32	31	73.8	99	2 C46518	Ig L1 chain V regi
33	31	73.8	115	2 B46518	Ig L1 chain V regi
34	31	73.8	171	2 AC2539	hypothetical prote
35	31	73.8	196	2 G84513	hypothetical prote
36	31	73.8	223	2 A91083	hypothetical prote
37	31	73.8	223	2 E65059	hypothetical prote
38	31	73.8	223	2 C85928	hypothetical prote
39	31	73.8	223	2 AH0409	conserved hypotnet
40	31	73.8	223	2 AB0859	conserved hypotnet
41	31	73.8	238	2 AB6562	hypothetical prote
42	31	73.8	238	2 C72061	conserved hypotnet
43	31	73.8	238	2 D71513	hypothetical prote
44	31	73.8	238	2 AB1671	conserved hypotnet
45	31	73.8	247	2 AG0250	conserved hypotnet

ALIGNMENTS

RESULT 1

S42466

Ig kappa chain V region - mouse

C:Species: Mus musculus (house mouse)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jan-2000

C:Accession: S42466

R.Shyanov, P.A.; Bessalov, I.A.; Terletskaya, H.N.; Deyev, S.M.

submitted to the EMBL Data Library, March 1994

A:Reference number: S42466

A:Accession: S42466

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-117 <SH1>

A:Cross-references: EMBL:X78108; NID:9460824; PIDN:CA54998.1; PID:9460825

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

C:Keywords: heterotetramer; Immunoglobulin

F:26-100/Domain: Immunoglobulin homology <IMM>

Query Match 100.0%; Score 42; DB 2; Length 117;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 1 WASTRHT 7

Db 60 WASTRHT 66

RESULT 2

P10207

anti-idiotypic antibody E225, kappa chain V region precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 21-Jan-2000

C:Accession: P10207

R.Souchon, H.; Doyen, N.; Riottot, M.M.; Rougeon, F.; Poljak, R.J.

MOL. Immunol. 27 429-433, 1990

A:Title: Nucleotide sequence of the VH, VL regions of an anti-idiotypic antibody rea

A:Reference number: P10207; MUID:90309764

A:Accession: P10207

A:Molecule type: mRNA

A:Residues: 1-131 <S0U>

A>Note: This sequence corresponds to subgroup V mouse immunoglobulin light chain

C:Superfamily: Immunoglobulin V region; Immunoglobulin homology

F:1-20/Domain: signal sequence #status predicted <SIG>

F:36-110/Domain: immunoglobulin homology <IMM>

F:44-54/Region: complementarity-determining 1

F:70-76/Region: complementarity-determining 2

F:109-117/Region: complementarity-determining 3

F:116-127/Region: J region

Query Match 100.0%; Score 42; DB 2; Length 131;

Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
|||||  
Db 70 WASTRHT 76

RESULT 3  
S30751

Ig kappa chain precursor V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 21-Jan-2000  
C:Accession: S30751  
R:Grant, F.J.; Levin, S.D.; Gilbert, T.; Kindsvogel, W.  
Nucleic Acids Res. 15, 5496, 1987  
A:Title: Improved RNA sequencing method to determine immunoglobulin mRNA sequence.  
A:Reference number: S30751; MUID:87260030  
A:Accession: S30751  
A:Molecule type: mRNA  
A:Residues: 1-152 <GRA>  
A:Cross-references: EMBL:X05877; NID:952195; PIDN:CAA29301.1; PID:952196  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:45-119/Domain: immunoglobulin homology <IMM>  
F:136-132/Domain: C region (C-kappa) (fragment) #status predicted <CRE>

Query Match 100.0%; Score 42; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 0.17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
|||||  
Db 79 WASTRHT 85

RESULT 4  
T27834

hypothetical protein ZK384.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T27834  
R:Almscough, R  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20427  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Accession: T27834  
A:Molecule type: DNA  
A:Residues: 1-212 <WIL>  
A:Cross-references: EMBL:Z82092; PIDN:CA805011.1; GSPDB:GN00023; CESP:ZK384.2  
C:Genetics:  
A:Gene: CESP:ZK384.2  
A:Map position: 5  
A:Insertions: 60/3; 106/3; 180/2  
C:Superfamily: yellowjacket venom allergen antigen 5

Query Match 81.0%; Score 34; DB 2; Length 212;  
Best Local Similarity 83.3%; Pred. No. 11;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
|||||  
Db 145 WASTRH 150

RESULT 5  
C38113

yebc protein - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Jun-2000

C:Accession: C38113; S19014; H64948  
R:Yatahagi, M.; Iwasaki, H.; Nakata, A.; Shinagawa, H.

J. Bacteriol. 173, 5747-5753, 1991

A:Title: Molecular analysis of the Escherichia coli ruvC gene, which encodes a Hollid

A:Reference number: A38113; MUID:91358366

A:Accession: C38113

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-246 <TA>

A:Cross-references: GB:D10165; GB:D90392; NID:9216649; PIDN:BAA01031.1; PID:9216652

R:Sharples, G.J.; Lloyd, R.G.

J. Bacteriol. 173, 7711-7715, 1991

A:Title: Resolution of Holliday junctions in Escherichia coli: identification of the

A:Reference number: S19013; MUID:92041688

A:Accession: S19014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-246 <SHA>

A:Cross-references: EMBL:X59551; NID:942172; PIDN:CAA42126.1; PID:942173

A:Experimental source: strain K-12

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H64948

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-246 <BLAT>

A:Cross-references: GB:AE000280; GB:U00099; NID:91788163; PIDN:AAC74934.1; PID:917881

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yebC

C:Superfamily: hypothetical protein MG332

Query Match 81.0%; Score 34; DB 2; Length 246;

Best Local Similarity 83.3%; Pred. No. 13;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
|||||  
Db 7 WASTRH 12

RESULT 6  
F90950

hypothetical protein ECs2574 [Imported] - Escherichia coli (strain O157:H7, substrain  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: F90950  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F90950  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA835997.1; PID:913362042; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 0509952  
C:Genetics:  
A:Gene: ECs2574  
C:Superfamily: hypothetical protein MG332

Query Match 81.0%; Score 34; DB 2; Length 246;

Best Local Similarity 83.3%; Pred. No. 13;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
|||||



Db 7 WANTRH 12

RESULT 7  
B85799  
hypothetical protein yebc [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: B85799  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: B85799  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <STO>  
A:Cross-references: GB:AE005174; MID:g12515920; PIDN:AG56854.1; GSPDB:GN00145; UWGP:Z29  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: yebc  
C:Superfamily: hypothetical protein MG332

Query Match 81.0%; Score 34; DB 2; Length 246;  
Best Local Similarity 83.3%; Pred. No. 13;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
||:||||  
Db 7 WANTRH 12

RESULT 8  
AG0743  
conserved hypothetical protein STY2107 [imported] - Salmonella enterica subsp. enterica  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AG0743  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
Th, T.; Connelton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AG0743  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD05650.1; PID:g16503146; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY2107  
C:Superfamily: hypothetical protein MG332

Query Match 81.0%; Score 34; DB 2; Length 246;  
Best Local Similarity 83.3%; Pred. No. 13;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
||:||||  
Db 7 WANTRH 12

RESULT 9  
T49780  
related to beta transducin-like protein [imported] - Neurospora crassa  
N:Alternate names: protein B9J10.170  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T49780  
R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Farthmann, B.; Holland, R.; Nykate  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49780  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-605 <SCH>  
A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.170  
A:Experimental source: BAC clone B9J10, strain OR74A  
C:Genetics:  
A:Gene: NCSP:B9J10.170  
A:Map position: 6  
A:Introns: 183/2

Query Match 81.0%; Score 34; DB 2; Length 605;  
Best Local Similarity 71.4%; Pred. No. 33;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
||:||||  
Db 201 WASTRHT 207

RESULT 10  
GNMWVW  
genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstr  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40244  
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 188, 102-113, 1992  
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping  
A:Reference number: A40244; MUID:92230206  
A:Accession: A40244  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <CHE>  
A:Cross-references: GB:M84754  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: Arg; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc  
E:1-115/Product: capsid protein C #status predicted <PC>  
E:116-101/Product: envelope protein M #status predicted <PM>  
E:192-389/Product: major envelope protein E #status predicted <ME>  
E:730-729/Product: nonstructural protein NS1 #status predicted <NS1>  
E:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
E:1007-1615/Product: hepatitisin #status predicted <NS3>  
E:1230-1237/Region: nucleotide-binding motif A (P-loop)  
E:1312-1317/Region: nucleotide-binding motif B  
E:1316-1319/Region: DEXH motif  
E:1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NA5>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
E:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041

Query Match 81.0%; Score 34; DB 1; Length 3010;  
Best Local Similarity 71.4%; Pred. No. 1,8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
||:||||  
Db 2816 WASTRHT 2822

RESULT 11  
C83590  
probable transcription regulator PA0436 [imported] - Pseudomonas aeruginosa (strain  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: C63590  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: C63590  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-206 <STO>  
A:Cross-references: GB:AE004481; GB:AE004091; NID:g9946293; PIDN:AA03825.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0436

Query Match 78.6%; Score 33; DB 2; Length 206;  
Best Local Similarity 83.3%; Pred. No. 17;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
||||:|  
Db 160 WASTQH 165

RESULT 12  
C90786  
probable tet operon regulator [imported] - *Escherichia coli* (strain O157:H7, substrain F  
C:Species: *Escherichia coli*  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C90786  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90786  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-212 <HAV>  
A:Cross-references: GB:BA000007; PIDN:BA834682.1; PID:g13360719; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: EC81259

Query Match 78.6%; Score 33; DB 2; Length 212;  
Best Local Similarity 83.3%; Pred. No. 18;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
||||:|  
Db 167 WASTQH 172

RESULT 13  
C64843  
probable transcription regulator ycdC - *Escherichia coli*  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C:Accession: C64843  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.U.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: C64843  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-212 <BLAT>  
A:Cross-references: GB:AE000203; GB:000096; NID:g1787248; PIDN:AA074098.1; PID:g1787249;  
A:Experimental source: strain K-12, substrain MG1655

C:Genetics:  
A:Gene: ycdC  
A:Keywords: DNA binding; transcription regulation  
F.39-58/Region: helix-turn-helix motif

Query Match 78.6%; Score 33; DB 2; Length 212;  
Best Local Similarity 83.3%; Pred. No. 18;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
||||:|  
Db 167 WASTQH 172

RESULT 14  
A85646  
probable tet operon regulator ycdC [imported] - *Escherichia coli* (strain O157:H7, sub  
C:Species: *Escherichia coli*  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: A85646  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206351  
A:Accession: A85646  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-212 <STO>  
A:Cross-references: GB:AE005174; NID:g12514373; PIDN:AA055629.1; GSPDB:GN00145; UMGF:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ycdC

Query Match 78.6%; Score 33; DB 2; Length 212;  
Best Local Similarity 83.3%; Pred. No. 18;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
||||:|  
Db 167 WASTQH 172

RESULT 15  
D97525  
hypothetical protein AGR\_C\_2525 [imported] - *Agrobacterium tumefaciens* (strain C58, C  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: D97525  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*  
A:Reference number: A87359; PMID:11743194  
A:Accession: D97525  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-275 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK67157.1; PID:g15156429; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_2525  
A:Map position: circular chromosome

Query Match 78.6%; Score 33; DB 2; Length 275;  
Best Local Similarity 71.4%; Pred. No. 24;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 7  
||||:|  
Db 212 WASTRH 218

Tue Aug 20 12:58:55 2002

us-09-824-286-4\_copy\_50\_56.rpr

Page 5

Search completed: August 20, 2002, 11:35:18  
Job time: 397 sec

2002 AUG 20 12:58:55

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:41:20 ; Search time 15.95 Seconds

(without alignments)  
16.993 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_50\_56

Perfect score: 42

Sequence: 1 WASTRHT 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	81.0	246	1 YEBE.ECOLI	P24237 escherichia
2	34	81.0	3010	1 POLG.HCVTW	P29846 h genome po
3	33	78.6	212	1 YCDC.ECOLI	P75899 escherichia
4	33	78.6	380	1 YC9B.SCHPO	O09885 schizosacch
5	33	78.6	1207	1 YL08.CAEEL	P34402 caenorhabdi
6	32	76.2	572	1 POXB.ECOLI	P07003 escherichia
7	32	76.2	889	1 TRK2.YEAST	P28384 saccharomyc
8	32	76.2	1330	1 VCAP.PRVIS	O00705 pseudorhabe
9	31	73.8	238	1 Y457.CHLTR	O84463 chlamydia t
10	31	73.8	238	1 Y573.CHLPN	O92730 chlamydia p
11	31	73.8	238	1 Y742.CHLMU	O9P155 chlamydia m
12	31	73.8	251	1 Y003.MYCLE	O49645 mycobacteri
13	31	73.8	251	1 Y003.MYCLE	O33214 mycobacteri
14	31	73.8	445	1 SCRC.RABIT	O46502 oryctolagus
15	31	73.8	449	1 SCRC.RABIT	P23811 rattus norv
16	31	73.8	461	1 PPBA.BACSU	O53241 mycobacteri
17	31	73.8	461	1 PPBA.BACSU	O53241 mycobacteri
18	31	73.8	502	1 SYE.MYCLE	O33120 mycobacteri
19	31	73.8	994	1 SYV.XYLFU	O9P155 chlamydia fas
20	31	73.8	1228	1 C1BA.BACTK	P05517 bacillus th
21	31	73.8	114	1 KV4A.HUMAN	P01655 homo sapien
22	30	71.4	121	1 KV40.HUMAN	P06312 homo sapien
23	30	71.4	133	1 KV4B.HUMAN	P06313 homo sapien
24	30	71.4	134	1 KV4C.HUMAN	P06313 homo sapien
25	30	71.4	134	1 KV4D.HUMAN	P06313 homo sapien
26	30	71.4	134	1 KV4E.HUMAN	P06313 homo sapien
27	30	71.4	134	1 KV4F.HUMAN	P06313 homo sapien
28	30	71.4	134	1 KV4G.HUMAN	P06313 homo sapien
29	30	71.4	134	1 KV4H.HUMAN	P06313 homo sapien
30	30	71.4	134	1 KV4I.HUMAN	P06313 homo sapien
31	30	71.4	134	1 KV4J.HUMAN	P06313 homo sapien
32	30	71.4	134	1 KV4K.HUMAN	P06313 homo sapien
33	30	71.4	134	1 KV4L.HUMAN	P06313 homo sapien

34	29	69.0	166	1 PPPI.PYRFU	O51732 pyrococcus
35	29	69.0	181	1 ORN.VIBCH	O9KX17 vibrio chol
36	29	69.0	232	1 FKB1.ABATF	O9LW17 arabidopsis
37	29	69.0	259	1 DEOC.VIBCH	O9KX17 vibrio chol
38	29	69.0	263	1 CTR2.CANFA	P04813 canis famli
39	29	69.0	342	1 GALR.SALTY	P74866 salmonella
40	29	69.0	343	1 GALR.ECOLI	P03024 escherichia
41	29	69.0	387	1 DADR.MOUSE	P51436 mus musculu
42	29	69.0	387	1 DADR.MOUSE	P30729 rattus norv
43	29	69.0	392	1 DADR.RAT	O93229 rana japoni
44	29	69.0	420	1 O63A.DROME	O9VZV8 drosophila
45	29	69.0	488	1 YGM9.YEAST	O01163 saccharomyc

#### ALIGNMENTS

RESULT ID	YEBE.ECOLI	STANDARD	PRT	246 AA.
AC	P24237			
DT	01-MAR-1992 (Rel. 21, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Protein yebc.			
GN	YEBE OR B1864.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91358366; PubMed=1885548;			
RA	Takahagi M., Iwasaki H., Nakata A., Shingawa H.,			
RT	"Molecular analysis of the Escherichia coli rucC gene, which encodes			
RT	a Holliday junction-specific endonuclease.";			
RT	J. Bacteriol. 173:5747-5753(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91358366; PubMed=1885548;			
RA	Takahagi M., Iwasaki H., Nakata A., Shingawa H.,			
RT	"Molecular analysis of the Escherichia coli rucC gene, which encodes			
RT	a Holliday junction-specific endonuclease.";			
RT	J. Bacteriol. 173:5747-5753(1991).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655.			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RT	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RT	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RT	Mau B., Siao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RT	Science 277:1453-1474(1997).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=K12;			
RX	MEDLINE=97251358; PubMed=9097040;			
RA	Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,			
RT	Isono K., Kasai H., Kimura S., Kitagawa M., Kitagawa M.,			
RT	Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,			
RT	Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,			
RT	Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,			
RT	Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.,			
RT	"A 460-kb DNA sequence of the Escherichia coli K-12 genome			
RT	corresponding to the 40.1-50.0 min region on the linkage map.";			
RT	DNA Res. 3:379-392(1996).			
RN	[5]			
RP	IDENTIFICATION BY MASS SPECTROMETRY.			
RX	MEDLINE=99420866; PubMed=10493123;			
RA	Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;			
RT	"Enrichment of low abundance proteins of Escherichia coli by			
RT	hydroxyapatite chromatography.";			



FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D53CDEF215 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 3010;  
 Best Local Similarity 71.4%; Pred. No. 78;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRHT 7  
 ||: |||  
 Db 2816 WASTRHT 2822

RESULT 3  
 YCOLI ECOLI STANDARD; PRT; 212 AA.  
 AC P75899;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical transcriptional regulator ycdC.  
 GN YCD C OR B1013.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NC NCB1\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 mln region on the linkage map.";  
 RL DNA Res. 3:137-155(1996)  
 CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000203; AAC74098.1; -  
 DR EMBL: D90738; BAA35790.1; -  
 DR Ecogene; EGI2301; ycdC.  
 DR InterPro; IPR001647; HTH\_Tetr.  
 DR Pfam; PF00440; tetr. 1.  
 DR PROSITE; PS01081; HTH\_TETR\_FAMILY; FALSE\_NEG.  
 KM Hypothetical protein; Transcription regulation; DNA-binding;  
 FT DNA\_BIND 39 58 H-T-H MOTIF (POTENTIAL).  
 SQ SEQUENCE 212 AA; 23687 MW; E2459B85DFAC277A CRC64;

Query Match 78.6%; Score 33; DB 1; Length 212;  
 Best Local Similarity 83.3%; Pred. No. 8.1;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WASTRH 6  
 ||||: |  
 Db 167 WASTRH 172

RESULT 4  
 YC9B\_SCHPO STANDARD; PRT; 380 AA.  
 AC O09885;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hypothetical 43.0 kDa protein C584.11C in chromosome III.  
 GN SPCC584.11C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 NC NCB1\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL032824; CAB37424.1; -  
 DR Hypothetical protein.  
 KM SEQUENCE 380 AA; A65777B672310E68 CRC64;  
 SQ SEQUENCE 380 AA; A65777B672310E68 CRC64;

Query Match 78.6%; Score 33; DB 1; Length 380;  
 Best Local Similarity 71.4%; Pred. No. 15;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRHT 7  
 ||||: |  
 Db 178 WASTRHT 184

RESULT 5  
 YL08\_CAEEL STANDARD; PRT; 1207 AA.  
 AC P34402;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 138.6 kDa protein F10E9.8 in chromosome III.  
 GN F10E9.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copley T., Cooper J., Coulson A.,  
 RA Craxton J., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Woldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----  
DR EMBL: L10986; AAA28012.2; -  
DR Wormpep: F10E9.8; CE00557.  
KW Hypothetical protein.  
SQ SEQUENCE 1207 AA; 138631 MW; EB13FBC52AC0A3FC CRC64;

Query Match 78.6%; Score 33; DB 1; Length 1207;  
Best Local Similarity 83.3%; Pred. No. 49;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
-----  
QY 1 WASTRH 6  
| | | | |  
DB 164 WGSTRH 169

RESULT 6  
POXB\_ECOLI STANDARD; PRT; 572 AA.  
ID POXB\_ECOLI  
AC P07003; Q47513; Q47514; Q47515; Q47516; Q47517; Q47518; Q47519;  
AC Q47520;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pyruvate dehydrogenase [cytochrome] (EC 1.2.2.2) (Pyruvate oxidase)  
DE (POX) (Pyruvate dehydrogenase [Ubiquinone]).  
GN POXB OR B0871.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-K12;  
RX MEDLINE-86286555; PubMed-3016647;  
RA Grabau C., Cronan J.E. Jr.;  
RT "Nucleotide sequence and deduced amino acid sequence of Escherichia  
RT coli pyruvate oxidase, a lipid-activated flavoprotein.";  
RL Nucleic Acids Res. 14:5449-5460(1986).  
-----  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Borna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
-----  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-97061202; PubMed-8905232;  
DR Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,

RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horichi T.;  
RT "A 718-bp DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
-----  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC STRAIN-K12;  
RX MEDLINE-89308683; PubMed-2663858;  
RA Grabau C., Chang Y.Y., Cronan J.E. Jr.;  
RT "Lipid binding by Escherichia coli pyruvate oxidase is disrupted by  
RT small alterations of the carboxyl-terminal region.";  
RL J. Biol. Chem. 264:12510-12519(1989).  
-----  
RP SEQUENCE OF 550-572 FROM N.A.  
RX MEDLINE-86033917; PubMed-3902830;  
RA Rechy M.A., Grabau C., Cronan J.E. Jr., Hager L.P.;  
RT "Characterization of the alpha-peptide released upon protease  
RT activation of pyruvate oxidase.";  
RL J. Biol. Chem. 260:14287-14291(1985).  
-----  
RP SEQUENCE OF 1-22 FROM N.A.  
RX MEDLINE-94293772; PubMed-8022274;  
RA Chang Y.Y., Wang A.Y., Cronan J.E. Jr.;  
RT "Expression of Escherichia coli pyruvate oxidase (POXB) depends on  
RT the sigma factor encoded by the rpoS(katF) gene.";  
RL Mol. Microbiol. 11:1019-1028(1994).  
-----  
CC -1- CATALYTIC ACTIVITY: Pyruvate + ferricytochrome b1 + H(2)O = CO(2)  
CC + acetate + ferrocyclochrome b1.  
CC -1- COFACTOR: THIAMINE PYROPHOSPHATE, FAD AND MAGNESIUM.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- PTM: ACTIVATED BY LIMITED PROTEOLYTIC DIGESTION. THIS CLEAVAGE  
CC PRODUCES A PEPTIDE (ALPHA-PEPTIDE) AND MIMICS THE ACTIVATION OF  
CC ENZYME BY PHOSPHOLIPIDS. THE PROTEOLYTIC CLEAVAGE ALSO RESULTS IN  
CC THE LOSS OF THE HIGH AFFINITY LIPID-BINDING SITE OF THE ENZYME.  
-----  
CC -1- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----  
DR EMBL: X04105; CAA27725.1; -  
DR EMBL: AE000186; AAC73958.1; -  
DR EMBL: D90724; BAA35585.1; -  
DR EMBL: S73268; AAB31180.1; -  
DR EMBL: M28208; AAB59101.1; -  
DR EMBL: L47688; AAB59102.1; -  
DR EMBL: L47689; AAB59103.1; -  
DR EMBL: L47690; AAB59104.1; -  
DR EMBL: L47691; AAB59105.1; -  
DR EMBL: L47692; AAB59106.1; -  
DR EMBL: L47693; AAB59107.1; -  
DR EMBL: L47694; AAB59108.1; -  
DR EMBL: L47695; AAB59109.1; -  
DR PIR: A23648; DEECPC.  
DR HSSP: P37063; IPOM.  
DR SWISS-2DPAGE: P07003; COLI.  
DR ECO2DBASE: G058.0; 6TH EDITION.  
DR Ecocode: EG10754; poxb.  
DR InterPro: IPR000399; TPP\_enzyme.  
DR Pfam: PF00205; TPP\_enzymes; 1.  
DR Pfam: PF02775; TPP\_enzymes-C; 1.  
DR Pfam: PF02776; TPP\_enzymes-N; 1.  
DR PROSITE: PS00187; TPP\_ENZYMES; 1.  
KW Oxidoreductase; Flavoprotein; FAD; Thiamine pyrophosphate; Magnesium;



KW Membrane; Lipid-binding: Complete proteome.  
 FT PEPTIDE 550 572 ALPHA-PEPTIDE.  
 FT ACT\_SITE 50 50 BY SIMILARITY.  
 FT MUTAGEN 533 533 A->T: IN POXB11.  
 FT MUTAGEN 553 553 A->V: IN POXB14.  
 FT MUTAGEN 560 560 D->P: IN POXB15; NORMAL ACTIVITY.  
 FT MUTAGEN 564 564 E->P: IN POXB16; LOSS OF ACTIVITY.  
 FT MUTAGEN 572 572 R->G: IN POXB10; REDUCED ACTIVITY; MAY  
 FT MUTAGEN 549 572 INTERACT LESS WITH MEMBRANES.  
 FT MUTAGEN 564 572 MISSING: IN POXB6.  
 FT MUTAGEN 570 572 MISSING: IN POXB7.  
 FT CONFLICT 364 365 MISSING: IN POXB8.  
 FT CONFLICT 414 416 OO -> HE (IN REF. 4).  
 SO SEQUENCE 572 AA; 62011 MW; 57B3BBE3A92BDEA CRC64;

Query Match 76.2%; Score 32; DB 1; Length 572;  
 Best Local Similarity 83.3%; Pred. No. 36;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6  
 Db 44 WASTRH 49

RESULT 7  
 TRK2\_YEAST STANDARD; PRT; 889 AA.  
 ID TRK2\_YEAST  
 AC P28584;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Potassium transport protein, low-affinity.  
 GN TRK2 OR RPD2 OR YKR050W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 NC NCBL\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91304419; PubMed=2072919;  
 RA "Ko C.H., Gaber R.F.;  
 RT "TRK1 and TRK2 encode structurally related K+ transporters in  
 RT Saccharomyces cerevisiae".  
 RL Mol. Cell. Biol. 11:4266-4273(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Vissers S., Urrestarazu L.A., Jauniaux J.-C.;  
 RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.  
 CC -!- FUNCTION: THIS PROTEIN IS REQUIRED FOR LOW-AFFINITY POTASSIUM  
 CC TRANSPORT.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M65215; AAA35172.1;  
 DR EMBL: Z28275; CAA82128.1;  
 DR PIR: A41259; A41259.  
 DR SGD: S0001758; TRK2.  
 DR InterPro: IPR003445; TRKH.  
 DR Pfam: PF02386; TRKH; 1.  
 KW TRANSPORT; Transmembrane; Potassium transport; Glycoprotein.  
 FT TRANSMEM 52 73 POTENTIAL.  
 FT TRANSMEM 81 101 POTENTIAL.  
 FT TRANSMEM 110 130 POTENTIAL.  
 FT TRANSMEM 465 487 POTENTIAL.

FT TRANSMEM 500 521 POTENTIAL.  
 FT TRANSMEM 525 545 POTENTIAL.  
 FT TRANSMEM 549 569 POTENTIAL.  
 FT TRANSMEM 565 605 POTENTIAL.  
 FT TRANSMEM 610 630 POTENTIAL.  
 FT TRANSMEM 658 678 POTENTIAL.  
 FT TRANSMEM 744 764 POTENTIAL.  
 FT TRANSMEM 777 797 POTENTIAL.  
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 889 AA; 101086 MW; 4A41448C9560276C CRC64;

Query Match 76.2%; Score 32; DB 1; Length 889;  
 Best Local Similarity 83.3%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6  
 Db 484 WASTRH 489

RESULT 8  
 VCAP\_PRIVIS STANDARD; PRT; 1330 AA.  
 ID VCAP\_PRIVIS  
 AC Q00705;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major capsid protein (MCP) (MCP142).  
 OS Pseudorabies virus (strain Indiana S) (PRV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 NC NCBL\_TaxID=31522;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92024125; PubMed=1718089;  
 RA Yamada S., Imada T., Watanabe W., Honda Y., Nakajima-Iijima S.,  
 RA Shimizu Y., Sekikawa K.;  
 RT "Nucleotide sequence and transcriptional mapping of the major capsid  
 RT protein gene of pseudorabies virus".  
 RL Virology 185:56-66(1991).  
 CC -!- FUNCTION: MAJOR PROTEIN OF THE ICOSAEDRAL CAPSID.  
 CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES MAJOR CAPSID PROTEIN  
 CC FAMILY.  
 CC PIR: A40777; VCBRS5.  
 DR InterPro: IPR000912; Herpes\_MCP.  
 DR Pfam: PF03122; Herpes\_MCP; 1.  
 DR PRINTS: PR00235; HSVCAPSIDMCP.  
 KW Coat protein.  
 SO SEQUENCE 1330 AA; 145937 MW; 4E228145F773A522 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 1330;  
 Best Local Similarity 71.4%; Pred. No. 86;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 7  
 Db 1205 WASTRH 1211

RESULT 9  
 YGCF\_ECOLI STANDARD; PRT; 223 AA.  
 ID YGCF\_ECOLI  
 AC P55139;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

DE Hypothetical protein Y9CF.
CN Y9CF OR B2777 OR Z4089 OR ECS3633.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubura S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1 SIMILARITY: 70 H.INFLUENZAE H1189.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29580; AAC69287.1; -
DR EMBL: AE000361; AAC75819.1; -
DR EMBL: AE005505; AAG57887.1; -
DR EMBL: AP002562; BAB37056.1; -
DR EcoGene: EG13034; y9CF.
DR InterPro: IPR001989; Radical_activat.
DR Pfam: PF02143; Radical_activat. 1
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 223 AA; 25029 MW; A717AAFI8F2A5D70 CRC64;

```

```

AC 084463;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT457.
GN CT457.
OS Chlamydia trachomatis.
OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/JW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Marathe R., Fan J., Atavind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1 SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001319; AAC68057.1; -
DR InterPro: IPR002876; DUF28.
DR Pfam: PF01709; DUF28. 1.
DR ProDom: PD004323; DUF28; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 238 AA; 26563 MW; 28C001245C0D3DF6 CRC64;

```

Query Match 73.8%; Score 31; DB 1; Length 238;  
 Best Local Similarity 66.7%; Pred. No. 23;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MASTRH 6
DB 7 WANTKH 12

```

RESULT 11  
 Y573.CHLPN STANDARD; PRT; 238 AA.

AC 0927Y0; Q9J012;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein CP00573/CP0176/CPJ0573.  
 GN CP00573 OR CP0176 OR CPJ0573.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Utterback T., Berry K., Bass S.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CML029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
CC EMBL; AE001642; AAD18713.1; -  
CC EMBL; AE002178; AAF38050.1; -  
CC EMBL; AP002547; BAA98779.1; -  
CC TIGR; CP0176; -  
CC DR InterPro: IPR002876; DUF28.  
CC DR Pfam: PF01709; DUF28; 1.  
CC DR ProDom: PD004323; DUF28; 1.  
CC DR Hypothetical protein; Complete proteome.  
CC KW Hypothetical protein; Complete proteome.  
CC SEQUENCE 238 AA; 26472 MW; 68E4F10F3B5FF772 CRC64;  
SQ  
  
Query Match 73.8%; Score 31; DB 1; Length 238;  
Best Local Similarity 66.7%; Pred. NO. 23;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WASTRH 6  
Db 7 WANTKH 12  
  
RESULT 12  
Y742\_CHLMU STANDARD; PRT; 238 AA.  
ID Y742\_CHLMU  
AC G9PJT5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein TC0742.  
GN TC0742.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mohn / Nig9;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Baas S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
CC EMBL; AE002342; AAF39551.1; -  
CC TIGR; TC0742; -  
CC DR InterPro: IPR002876; DUF28.  
CC DR Pfam: PF01709; DUF28; 1.  
CC DR ProDom: PD004323; DUF28; 1.  
CC DR Hypothetical protein; Complete proteome.  
CC KW Hypothetical protein; Complete proteome.  
CC SEQUENCE 238 AA; 26486 MW; C5D5657B0909264C CRC64;  
SQ  
  
Query Match 73.8%; Score 31; DB 1; Length 238;  
Best Local Similarity 66.7%; Pred. NO. 23;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WASTRH 6  
Db 7 WANTKH 12  
  
RESULT 13  
Y003\_MYCLE STANDARD; PRT; 251 AA.  
ID Y003\_MYCLE  
AC Q49645;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 26.6 kDa protein ML0475.  
GN ML0475 OR U1177B OR B1177\_C2\_181.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Smith D.R., Robison K.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,  
RA Holtroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus.";  
RL Nature 403:1007-1011(2001).  
CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
CC EMBL; U00011; AAA17094.1; -  
CC EMBL; AL583918; CAC29983.1; -  
CC Leproma; ML0475; -  
CC DR InterPro: IPR002876; DUF28.  
CC DR Pfam: PF01709; DUF28; 1.  
CC DR ProDom: PD004323; DUF28; 1.  
CC DR Hypothetical protein; Complete proteome.  
CC KW Hypothetical protein; Complete proteome.  
CC SEQUENCE 251 AA; 26614 MW; 8584C11AE1F02E6C CRC64;  
SQ

Query Match 73.8%; Score 31; DB 1; Length 251;  
 Best Local Similarity 66.7%; Pred. No. 25;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WASTRH 6  
 11:1:1  
 DB 7 WATTKH 12

RESULT 14  
 Y003\_MYCTU STANDARD; PRT; 251 AA.  
 AC 033214;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 26.8 kDa protein RV2603C.  
 GN RV2603C OR MT2678 OR MTC1270A.02.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,  
 RA Badoev K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwyn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J.A., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z98045; CAB10838.1; -  
 DR EMBL: AE007101; AAK46994.1; -  
 DR TIGR: MT2678; -  
 DR Tuberculist: RV2603C; -  
 DR InterPro: IPR002876; DUF28.  
 DR Pfam: PF01709; DUF28.1.  
 DR ProDom: PD004323; DUF28.1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 251 AA; 26799 MW; B163DA3729A06B33 CRC64;

Query Match 73.8%; Score 31; DB 1; Length 251;  
 Best Local Similarity 66.7%; Pred. No. 25;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WASTRH 6  
 11:1:1  
 DB 7 WATTKH 12

RESULT 15  
 SCRC\_RABIT STANDARD; PRT; 445 AA.  
 AC 046502;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Secretin receptor precursor (SC-R).  
 GN SCRR.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98366112; PubMed=9700755;  
 RA Svoboda M., Tstenoy M., de Neef P., Delporte C., Waelbroeck M.,  
 RA Koberecht P.;  
 RT "Molecular cloning and in vitro properties of the recombinant rabbit  
 RT secretin receptor."  
 RL Peptides 19:1055-1062(1998).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR SECRETIN. THE ACTIVITY OF THIS  
 CC RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL  
 CC CYCLASE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF025411; AAC32767.1; -  
 DR InterPro: IPR000832; GPCR\_secretin.  
 DR InterPro: IPR001879; hormn\_receptor.  
 DR Pfam: PF00002; 7tm\_2; 1.  
 DR Pfam: PF02793; HRM; 1.  
 DR PRINTS: PR00249; GPCRSECRETIN.  
 DR SMART: SM00068; HormR; 1.  
 DR PROSITE: PS00649; G\_PROTEIN\_RECEP\_F2\_1; 1.  
 DR PROSITE: PS00650; G\_PROTEIN\_RECEP\_F2\_2; 1.  
 DR PROSITE: PS50227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE: PS50261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 445  
 FT DOMAIN 22 139  
 FT TRANSMEM 140 163  
 FT DOMAIN 164 170  
 FT TRANSMEM 171 190  
 FT DOMAIN 191 212  
 FT TRANSMEM 213 236  
 FT DOMAIN 237 250  
 FT TRANSMEM 251 272  
 FT DOMAIN 273 290  
 FT TRANSMEM 291 313  
 FT DOMAIN 314 339  
 FT TRANSMEM 340 358  
 FT DOMAIN 359 365  
 FT TRANSMEM 366 388  
 FT DOMAIN 389 445  
 FT CARBOHYD 68 68  
 FT CARBOHYD 96 96  
 FT N-LINKED (GLCNAC...) (POTENTIAL).  
 FT N-LINKED (GLCNAC...) (POTENTIAL).  
 FT N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 445 AA; 50495 MW; 31CA169CB099F194 CRC64;

Query Match 73.8%; Score 31; DB 1; Length 445;  
Best Local Similarity 83.3%; Pred. No. 44;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRH 6  
11111

Db 269 WAVTRH 274

Search completed: August 20, 2002, 11:41:21  
Job time: 400 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:40:53 ; Search time 43.57 Seconds  
(without alignments)  
27.794 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_50\_56

Perfect score: 42

Sequence: 1 WASTRHT 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPPREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.rvitus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	387	2 Q9R6M6	Q9R6M6 agrobacteri
2	34	81.0	199	10 Q9LVE8	Q9LVE8 arabisdopsi
3	34	81.0	212	5 Q62508	Q62508 caenorhabdi
4	34	81.0	435	10 Q9FMJ7	Q9FMJ7 arabisdopsi
5	34	81.0	605	3 Q9P564	Q9P564 neurospora
6	34	81.0	743	10 Q9LW44	Q9LW44 lycopersico
7	34	81.0	746	10 Q9LW43	Q9LW43 lycopersico
8	33	78.6	206	16 Q9LW43	Q9LW43 lycopersico
9	33	78.6	313	5 Q9S0L5	Q9S0L5 pseudomonas
10	33	78.6	347	16 Q9L5J7	Q9L5J7 pseudomonas
11	33	78.6	650	5 Q9S0L4	Q9S0L4 drosophila
12	32	76.2	132	4 Q9BUN6	Q9BUN6 homo sapien
13	32	76.2	142	16 Q9PC08	Q9PC08 xylella fas
14	32	76.2	222	2 Q9X5F1	Q9X5F1 zymomonas m
15	32	76.2	263	2 Q84969	Q84969 rhodococcus
16	32	76.2	268	16 Q06172	Q06172 mycobacteri

17	32	76.2	268	16 Q07709	Q07709 mycobacteri
18	32	76.2	337	4 Q9NZ76	Q9NZ76 homo sapien
19	32	76.2	345	4 Q9NZK9	Q9NZK9 homo sapien
20	32	76.2	434	16 Q92A01	Q92A01 listeria in
21	32	76.2	445	2 Q93J94	Q93J94 streptomyce
22	32	76.2	446	1 Q9V2V8	Q9V2V8 thermoprote
23	32	76.2	462	2 Q9FBI5	Q9FBI5 streptomyce
24	32	76.2	503	12 Q9J303	Q9J303 ectocarpus
25	32	76.2	672	5 Q9NK74	Q9NK74 drosophila
26	32	76.2	725	16 Q9RS62	Q9RS62 delnoccocus
27	32	76.2	738	10 Q48929	Q48929 nicotiana t
28	32	76.2	745	16 Q9RT19	Q9RT19 delnoccocus
29	32	76.2	818	5 Q9VJN0	Q9VJN0 drosophila
30	32	76.2	824	10 Q64776	Q64776 arabisdopsi
31	32	76.2	1330	12 Q04534	Q04534 pseudorabie
32	32	76.2	1385	12 Q65565	Q65565 bovine herp
33	32	76.2	1468	5 Q9V854	Q9V854 drosophila
34	32	76.2	2334	12 Q96725	Q96725 european br
35	31	73.8	127	4 Q96NH1	Q96NH1 homo sapien
36	31	73.8	196	10 Q9ZPT2	Q9ZPT2 arabisdopsi
37	31	73.8	198	2 Q9F1W0	Q9F1W0 streptomyce
38	31	73.8	237	2 Q9RJT0	Q9RJT0 streptomyce
39	31	73.8	250	2 Q9L288	Q9L288 streptomyce
40	31	73.8	251	2 Q9AE12	Q9AE12 cornebacte
41	31	73.8	288	2 Q51543	Q51543 pseudomonas
42	31	73.8	394	2 Q914X0	Q914X0 streptomyce
43	31	73.8	429	5 Q9N8Y2	Q9N8Y2 trypanosoma
44	31	73.8	448	2 Q93076	Q93076 rhodococcus
45	31	73.8	448	10 Q9XHG1	Q9XHG1 graciilaria

## ALIGNMENTS

RESULT	ID	Q9R6M6	PRELIMINARY:	PRT:	387 AA.
Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6
AC	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6	Q9R6M6
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	TIOF33	PROTEIN.			
GN	TIOF33				
OS	Agrobacterium tumefaciens.				
OC	Plasmid pTi-SAKURA.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Rhizobium.				
OX	NCBI_TaxID=358;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MAFF301001;				
RX	MEDLINE=20184752; PubMed=10721727;				
RA	Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,				
RA	Katoh A., Yoshida K.;				
RT	"Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";				
RL	Gene 242:331-336(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MAFF301001;				
RX	MEDLINE=98193120; PubMed=9524202;				
RA	Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;				
RT	"Novel structural difference between nopaline- and octopine- type ttrb				
RT	gene:construction of genetic and physical map and sequencing of				
RT	ttr/ttrai and rep gene clusters of a new Ti plasmid pTi-SAKURA.";				
RL	Biochim. Biophys. Acta 1396:1-7(1998).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MAFF301001;				
RA	Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;				
RT	"Genome structure of pTi-SAKURA(1): Strategy for DNA sequencing of a				
RT	Japanese cherry-Ti plasmid.";				
RL	Nucleic Acids Symp. Ser. 37:159-160(1998).				
RN	[4]				

RP SEQUENCE FROM N.A.  
RC STRAIN-MAFF301001;  
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;  
RT "Genome structure of pri-SAKURA (III): Characteristics of T-DNA.";   
RL Nucleic Acids Symp. Ser. 39:185-186(1998).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAFF301001;  
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;  
RT "Genome structure of pri-SAKURA (IV): Characteristics of tra region.";   
RN Nucleic Acids Symp. Ser. 39:187-188(1998).  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAFF301001;  
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;  
RT "Genome structure of pri-SAKURA (V): Complete nucleotide sequence of   
RT plasmid pri-SAKURA's vir region in Agrobacterium tumefaciens.";   
RL Nucleic Acids Symp. Ser. 39:265-266(1998).  
DR EMBL: AB016260; BAA87658.1; -.  
KM Plasmid.  
SQ SEQUENCE 387 AA; 43706 MW; 67AE002F5A6684F6 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 387;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
DB 120 WASTRH 125

RESULT 2  
ID O9LVE8 PRELIMINARY; PRT; 199 AA.  
AC O9LVE8;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE: M1L23.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
CX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-COLUMBIA;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-COLUMBIA;  
RX MEDLINE=20277480; PubMed=10819329;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence   
RT features of the regions of 4,504,864 bp covered by sixty P1 and YAC   
RT clones.";   
RL DNA Res. 7:131-135(2000).  
DR EMBL: AB019232; BAB02342.1; -.  
SQ SEQUENCE 199 AA; 22873 MW; 2ABE82C5727F7AE9 CRC64;

Query Match 81.0%; Score 34; DB 10; Length 199;  
Best Local Similarity 83.3%; Pred. No. 24;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
DB 107 WASTRH 112

RESULT 3

Query Match

O62508  
ID O62508 PRELIMINARY; PRT; 212 AA.  
AC O62508;  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE ZK384.2 PROTEIN.  
GN ZK384.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nemata; Chordata; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ainscough R.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for   
RT investigating biology.";   
RL Science 282:2012-2018(1998).  
DR EMBL: Z82092; CAB05011.1; -.  
DR InterPro: IPR001283; SCP. 2.  
DR Pfam: PF00188; SCP. 2.  
DR PRINTS: PR00837; V5TPXLIKE.  
DR SMART: SM00198; SCP. 1.  
SQ SEQUENCE 212 AA; 24122 MW; B25A6D727EAC3B7 CRC64;

Query Match 81.0%; Score 34; DB 5; Length 212;  
Best Local Similarity 83.3%; Pred. No. 25;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
DB 145 WASTRH 150

RESULT 4  
ID O9FMJ7 PRELIMINARY; PRT; 435 AA.  
AC O9FMJ7;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
DE SIMILARITY TO ZINC FINGER PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-COLUMBIA;  
RX MEDLINE=98162728; PubMed=9501997;  
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,   
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.   
RT Sequence features of the regions of 1,191,918 bp covered by seventeen   
RT physically assigned P1 clones.";   
RL DNA Res. 4:401-414(1997).  
DR EMBL: AB008265; BAB10568.1; -.  
DR InterPro: IPR000571; ZF-CCCH.  
DR Pfam: PF00642; ZF-CCCH. 5.  
DR SMART: SM00356; ZNF\_C3H1. 5.  
SQ SEQUENCE 435 AA; 48606 MW; 8BACBB780A03C6F2 CRC64;

Query Match 81.0%; Score 34; DB 10; Length 435;  
Best Local Similarity 83.3%; Pred. No. 55;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;



OY 1 WASTRHT 6  
1:||||  
Db 254 WSTRHT 259

RESULT 5  
ID 09P564 PRELIMINARY; PRT; 605 AA.  
AC 09P564;  
DT 01-OCT-2000 (TREMBLREL. 15, Created)  
DT 01-OCT-2000 (TREMBLREL. 15, Last sequence update)  
DE RELATED TO BETA TRANSDUCIN-LIKE PROTEIN.  
GN B9J10.170.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: ALJ56324; CAB92024.1; -  
SQ SEQUENCE 605 AA; 69676 MW; 1ED03092C7A894A3 CRC64;

Query Match 81.0%; Score 34; DB 3; Length 605;  
Best Local Similarity 71.4%; Pred. No. 78;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 WASTRHT 7  
1:||||  
Db 201 WASTRHT 207

RESULT 6  
ID 09LMA4 PRELIMINARY; PRT; 743 AA.  
AC 09LMA4;  
DT 01-OCT-2000 (TREMBLREL. 15, Created)  
DT 01-OCT-2000 (TREMBLREL. 15, Last sequence update)  
DE 01-DEC-2001 (TREMBLREL. 19, Last annotation update)  
DE SUBTILISIN-LIKE PROTEASE.  
GN P69E.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LEAF;  
RX MEDLINE=20098714; PubMed=10631250;  
RA Jorda L., Conejero V., Vera P.;  
RT "Characterization of two differentially regulated genes (P69E and  
RT P69F) encoding new members of the subtilisin-like protease clan from  
RT tomato plants.";  
RL Plant Physiol. 122:67-76(2000).  
DR EMBL: Y18931; CAB67119.1; -  
DR HSSP: Q99405; 1MPT.  
DR MEROPS: S08.006; -;  
DR InterPro: IPR003137; PA.  
DR InterPro: IPR00209; Peptidase\_S8.  
DR Pfam: PF02225; PA; 1.  
DR Pfam: PF00082; Peptidase\_S8; 3.  
DR PRINTS: PR00723; SUBTILISIN.  
DR PROSITE: PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
DR PROSITE: PS00136; SUBTILASE\_ASP; UNKNOWN\_1.

DR PROSITE: PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
KW Protease.  
SQ SEQUENCE 743 AA; 79134 MW; 63EE2F2B2510BCA CRC64;

Query Match 81.0%; Score 34; DB 10; Length 743;  
Best Local Similarity 71.4%; Pred. No. 97;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 WASTRHT 7  
1:||||  
Db 727 WSTRHTS 733

RESULT 7  
ID 09LMA3 PRELIMINARY; PRT; 746 AA.  
AC 09LMA3;  
DT 01-OCT-2000 (TREMBLREL. 15, Created)  
DT 01-OCT-2000 (TREMBLREL. 15, Last sequence update)  
DE 01-DEC-2001 (TREMBLREL. 19, Last annotation update)  
DE SUBTILISIN-LIKE PROTEASE.  
GN P69F.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LEAF;  
RX MEDLINE=20098714; PubMed=10631250;  
RA Jorda L., Conejero V., Vera P.;  
RT "Characterization of two differentially regulated genes (P69E and  
RT P69F) encoding new members of the subtilisin-like protease clan from  
RT tomato plants.";  
RL Plant Physiol. 122:67-76(2000).  
DR EMBL: Y18932; CAB67120.1; -  
DR HSSP: P29599; 1ST3.  
DR MEROPS: S08.006; -;  
DR InterPro: IPR003137; PA.  
DR InterPro: IPR00209; Peptidase\_S8.  
DR Pfam: PF02225; PA; 1.  
DR Pfam: PF00082; Peptidase\_S8; 3.  
DR PRINTS: PR00723; SUBTILISIN.  
DR PROSITE: PS00136; SUBTILASE\_ASP; UNKNOWN\_1.  
DR PROSITE: PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
KW Protease.  
SQ SEQUENCE 746 AA; 79062 MW; F2B9D6B7EA0E9A4B CRC64;

Query Match 81.0%; Score 34; DB 10; Length 746;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 WASTRHT 7  
1:||||  
Db 730 WSTRHTS 736

RESULT 8  
ID 09I681 PRELIMINARY; PRT; 206 AA.  
AC 09I681;  
DT 01-MAR-2001 (TREMBLREL. 16, Created)  
DT 01-MAR-2001 (TREMBLREL. 16, Last sequence update)  
DE 01-OCT-2001 (TREMBLREL. 18, Last annotation update)  
DE PROBABLE TRANSCRIPTIONAL REGULATOR.  
GN PA0436.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;

RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
REGULATORS.  
CC EMBL; AE004481; AAC03825.1; -  
DR Interpro: IPR001647; HTH\_Tetr.  
DR Pfam: PF00440; tetr. 1.  
DR PRINTS: PR00455; HTHTETR.  
KW Complete proteome: DNA-binding; Transcription regulation.  
SQ SEQUENCE 206 AA; 23718 MW; 365E298E435FE17B CRC64;

Query Match 78.6%; Score 33; DB 16; Length 206;  
Best Local Similarity 83.3%; Pred. No. 40;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WASTRH 6  
Db 160 WASTRH 165

RESULT 9  
O950L5 PRELIMINARY; PRT; 313 AA.  
AC O950L5;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHEICAL 36.3 KDA PROTEIN.  
GN F10E9.10.  
OS *Caenorhabditis elegans*.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderiinae; *Caenorhabditis*.  
OX NCBI\_TaxID=6239;  
RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode *C. elegans*: a platform for  
investigating biology. The *C. elegans* Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Anderson K.;  
RT "The sequence of *C. elegans* cosmid F10E9.";  
RL submitted (Feb-1993) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L10986; AAK93846.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 313 AA; 36253 MW; 9DC63002C30F9AF3 CRC64;

Query Match 78.6%; Score 33; DB 5; Length 313;  
Best Local Similarity 83.3%; Pred. No. 62;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRH 6  
Db 164 WASTRH 169

RESULT 10  
O915J7 PRELIMINARY; PRT; 347 AA.  
AC O915J7;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE HYPOTHEICAL PROTEIN PA0732.  
GN PA0732.  
OS *Pseudomonas aeruginosa*.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004508; AAC04121.1; -  
KW Hypothetical protein: Complete proteome.  
SQ SEQUENCE 347 AA; 37161 MW; 264E4DC064290CC3 CRC64;

Query Match 78.6%; Score 33; DB 16; Length 347;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 WASTRHT 7  
Db 318 WASTRHT 324

RESULT 11  
O950I4 PRELIMINARY; PRT; 650 AA.  
AC O950I4;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE GH14380P.  
GN CG12239.  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; *Drosophila*.  
OX NCBI\_TaxID=7227;  
RA [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phoumenavong S., Wan K.,  
Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY058378; AAL13607.1; -  
SQ SEQUENCE 650 AA; 80172 MW; 106796301C835D67 CRC64;

Query Match 78.6%, Score 33; DB 5; Length 650;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WASTRH 6  
| | | | |  
DB 22 WASSRH 27

## RESULT 12

Q9BUN6 PRELIMINARY; PRT; 132 AA.

AC Q9BUN6; 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE SIMILAR TO PROGRAMMED CELL DEATH 9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-EYE, RETINOBLASTOMA;  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: BC002460; AAH02460.1; -  
SQ SEQUENCE 132 AA; 14563 MW; 749ED93CF43D9A34 CRC64;

Query Match 76.2%; Score 32; DB 4; Length 132;  
Best Local Similarity 71.4%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRH 7  
| | | | |  
DB 85 WASSRH 91

## RESULT 13

Q9PG08 PRELIMINARY; PRT; 142 AA.

AC Q9PG08; 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHELICAL PROTEIN XF0240.  
GN XF0240.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
Buono M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,  
Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
Facinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
Ho P.L., Honaisel J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P.,  
Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,  
Marques M.V., Martins E.A.L., Martins E.M.F., Matsumura A.Y.,  
Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
Nhaui A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,  
Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
Ra Queiro R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen xylella fastidiosa.";  
RL Nature 406:151-159(2000).  
DR EMBL: AE003878; AAF83053.1; -  
DR InterPro: IPR002577; DUF24.  
DR Pfam: PF01638; DUF24; 1.  
DR ProDom: PD004032; DUF24; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 142 AA; 15878 MW; CAD3A2B6EDB13E31 CRC64;

Query Match 76.2%; Score 32; DB 16; Length 142;  
Best Local Similarity 83.3%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRH 6  
| | | | |  
DB 113 WASTRH 118

## RESULT 14

Q9X5F1 PRELIMINARY; PRT; 222 AA.

AC Q9X5F1; 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE HYPOTHELICAL 25.0 KDA PROTEIN.  
OS Zymomonas mobilis.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
OC Zymomonas.  
OX NCBI\_TaxID=542;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ZM4;  
RA Lee H.J., Kang H.S.;  
RT "Sequence analysis of 43D2 fosmid clone of Zymomonas mobilis ZM4.";  
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF124757; AAD29658.1; -  
DR InterPro: IPR001441; UPP\_synth.  
DR Pfam: PF01255; UPP\_synthetase; 1.  
DR ProDom: PD003461; UPP\_synth; 1.  
DR PROSITE: PS01066; UPP\_SYNTHETASE; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 222 AA; 24977 MW; EA9723DCB30F42FB CRC64;

Query Match 76.2%; Score 32; DB 2; Length 222;  
Best Local Similarity 83.3%; Pred. No. 69;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 WASTRH 6  
| | | | |  
DB 7 WASSRH 12

## RESULT 15

Q84989 PRELIMINARY; PRT; 263 AA.

AC Q84989; 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE PUTATIVE TRANSPOSASE (FRAGMENT).  
OS Rhodococcus opacus (Nocardia opaca).

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.  
OX NCBI\_TaxID-37919;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1CP;  
RX MEDLINE-98324954; PubMed-9657989;  
RA Seibert V., Kourbatova E.M., Golovleva L.A., Schlomann M.;  
RT "Characterization of the maleylacetate reductase MacA of Rhodococcus  
RT opacus 1CP and evidence for the presence of an isofunctional enzyme.";  
RL J. Bacteriol. 180:3503-3508 (1998).  
DR EMBL; AF030176; AAC38799.1; -;  
DR InterPro: IPR002525; Transposase\_9.  
DR Pfam; PF01548; Transposase\_9; 1.  
FT NON\_TER 263 263  
SQ SEQUENCE 263 AA; 28327 MM; 64DBAD92CB4317D9 CRC64;

Query Match 76.2%; Score 32; DB 2; Length 263;  
Best Local Similarity 71.4%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 WASTRHT 7  
1 |||||  
DB 151 WERTRHT 157

Search completed: August 20, 2002, 11:40:54  
Job time: 433 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:15:11 ; Search time 94.22 Seconds  
(Without alignments)  
11.789 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_95\_104  
Perfect score: 61  
Sequence: 1 PWTFCGCTKL 10

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_032802.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	19	21	AA801631
2	61	100.0	20	14	AA833254
3	61	100.0	104	13	AA830010
4	61	100.0	106	19	AA831647
5	61	100.0	107	12	AA811985
6	61	100.0	107	13	AA825411
7	61	100.0	107	14	AA838614
8	61	100.0	107	14	AA81283
9	61	100.0	107	19	AA866098
10	61	100.0	107	19	AA849812
11	61	100.0	107	19	AA858515

12	61	100.0	107	19	AA858505	H65 protein sequen
13	61	100.0	107	19	AA858478	Humanised anti-CD5
14	61	100.0	108	13	AA821292	Murine VL kappa gr
15	61	100.0	108	13	AA821310	Light chain of M16
16	61	100.0	108	16	AA879160	Human I9E receptor
17	61	100.0	108	17	AA800834	Variable light cha
18	61	100.0	108	17	AA804177	Variant variable 1
19	61	100.0	108	17	AA800828	Variable light cha
20	61	100.0	108	18	AA816620	Anti-human FasL an
21	61	100.0	108	18	AA827360	Light chain variab
22	61	100.0	108	18	AA827355	Light chain variab
23	61	100.0	108	19	AA807620	Anti-VEGF murine a
24	61	100.0	108	22	AA807484	Synthetic antibody
25	61	100.0	108	22	AA807486	Synthetic antibody
26	61	100.0	108	22	AA807488	Synthetic antibody
27	61	100.0	108	22	AA807480	Synthetic antibody
28	61	100.0	108	22	AA807452	Synthetic antibody
29	61	100.0	108	22	AA807454	Synthetic antibody
30	61	100.0	108	22	AA807496	Synthetic antibody
31	61	100.0	108	22	AA807517	Antibody scFv8 11g
32	61	100.0	108	22	AA881980	Ganglioside GD3 sp
33	61	100.0	109	16	AA879858	Anti-EGFR antibody
34	61	100.0	109	18	AA818271	Prp 37 light chain
35	61	100.0	109	22	AA85910	Anti-Prp antibody
36	61	100.0	109	22	AA85862	Light chain variab
37	61	100.0	110	13	AA830012	Light chain variab
38	61	100.0	109	22	AA860810	Murine anti-HIV mu5
39	61	100.0	111	15	AA855123	Anti HIV antibody
40	61	100.0	111	15	AA860302	Murine M195 antibo
41	61	100.0	111	22	AA869669	VI region from an
42	61	100.0	112	16	AA880272	Variable light cha
43	61	100.0	112	17	AA800830	Murine 5D12 monocl
44	61	100.0	112	18	AA823024	Humanized 5D12 mon
45	61	100.0	112	18	AA823026	

## ALIGNMENTS

RESULT 1  
ID AAB01631 standard; Peptide: 19 AA.  
XX AAB01631;  
XX  
DT 07-DEC-2000 (first entry)  
XX  
DE Chimaeric immunoglobulin ligation sequence #3.  
XX  
KW Immunoglobulin: variable region; Light chain; L chain; cancer;  
humanised antibody.  
XX  
OS Chimeric - Mus sp.  
OS Chimeric - Homo sapiens.  
XX  
PN EPI013761-A2.  
XX  
PD 28-JUN-2000.  
XX  
PF 18-SEP-1992; 99EP-0124345.  
XX  
PR 18-SEP-1991; 91JP-0238375.  
PR 18-SEP-1992; 92EP-0116026.  
XX  
PA (KYOW) KYOWA HAKKO KOGYO KK.  
XX  
PI Shitara K, Hanai N, Hasegawa M, Miyaji H, Kuwana Y;  
XX WPI: 2000-402204/35.  
DR N-PSDB: AAA51007.  
XX  
PT New humanized chimera antibody KM-871 useful for treating cancer,  
comprises variable region of mouse monoclonal antibody, reactive with

PT ganglioside and human antibody constant region -  
XX  
PS Example 2; Page 29; 65pp; English.  
XX  
CC The present sequence is a ligation sequence used to join the protein  
CC sequence for a murine immunoglobulin light chain variable region from  
CC plasmid KM-641 with a human light chain antibody protein sequence. This  
CC was used in an expression vector to produce humanised chimaeric  
CC antibodies, which can be used to treat cancer. Humanised chimaeric  
CC antibodies are more effective than mouse antibodies as they do not  
CC provoke a reaction in the human and side effects, such as the formation  
CC of anti-mouse immunoglobulin antibody and the rapid half-life of the  
CC immunoglobulins, do not occur.  
XX  
SQ Sequence 19 AA;

Query Match 100.0%; Score 61; DB 21; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGTRL 10  
|||  
Db 2 pwtfggtrl 11

RESULT 2  
AAR33254  
ID AAR33254 standard; Protein: 20 AA.  
XX  
AC AAR33254;

XX 12-JUL-1993 (first entry)

DE PKM641A2 derived L chain linker.

XX Linker; humanised; chimeric; antibody; expression vector; heavy;  
KW light; chain; variable region.

XX Synthetic.

XX EP53199-A.

PD 24-MAR-1993.

PF 18-SEP-1992; 92EP-0116026.

PR 18-SEP-1991; 91JP-0238375.

XX (KYOW ) KYOWA HAKKO KOGYO CO LTD.

XX Hanai N, Hasegawa M, Kuwana Y, Miyaji H, Shitara K;

XX WPI: 1993-095510/12.

DR N-PSDB; AAQ37055.

XX Humanised chimeric antibody prodn. against ganglioside GD3 - for  
PT treating cancers, such as melanoma, neuroblastoma, etc.

XX Claim 15; Page 31; 63pp; English.

XX The sequences given in AAR33253-54 are linkers which were used in the  
CC construction of humanised chimeric antibodies. The linkers are  
CC used in construction of heavy and light chain variable region  
CC antibodies respectively.

XX Sequence 20 AA;

Query Match 100.0%; Score 61; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.0071;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGTRL 10  
|||  
Db 2 pwtfggtrl 11

RESULT 3  
AAR30010  
ID AAR30010 standard; Protein: 104 AA.  
XX  
AC AAR30010;

XX 29-APR-1993 (first entry)

DE Light chain variable region of anti-p24 antibody.

XX MAb: monoclonal antibody; AIDS; HIV; infection; disease; immunoassay;  
KW p24; antibody-dependent cell-mediated cytotoxicity; macrophages;  
KW lymphoid cells.

OS Chimaeric; Homo sapiens; Mus musculus.

PN EP519866-A.

PD 23-DEC-1992.

PF 10-JUN-1993; 93EP-0810445.

PR 18-JUN-1991; 91EP-0810468.

XX (CIBA ) CIBA GEIGY AG.

XX MESTAN J, LAZDINS JK, WOODS-COOK KA, HARDMAN N, HOCHKEPPEL H;

XX WPI: 1992-426048/52.

XX Monoclonal antibody specific for HIV core protein p24 - mediate  
PT antibody-dependent cell-mediated cytotoxicity, for treating,  
PT preventing and diagnosing HIV infection.

XX Example 4; Page 22; 44pp; English.

XX This sequence represents the light chain variable region from  
CC anti-p24 antibody from hybridoma 25-57-1. The MAb is used to  
CC prevent progression of AIDS and to treat HIV infections, and  
CC also as an immunoassay reagent for qualitative or quantitative  
CC analysis of p24 for diagnostic purposes.

XX Sequence 104 AA;

Query Match 100.0%; Score 61; DB 13; Length 104;  
Best Local Similarity 100.0%; Pred. No. 0.034;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGTRL 10  
|||  
Db 93 pwtfggtrl 102

RESULT 4  
AAW31647  
ID AAW31647 standard; Protein: 106 AA.  
XX  
AC AAW31647;

XX 21-MAY-1998 (first entry)

DE Monoclonal antibody CP.B8 light chain variable region.

XX Cytokine receptor; gamma common chain; gc chain; human;  
KW blocking agent; monoclonal antibody; CP.B8; immunological disease;;  
KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;  
KW insulin-dependent diabetes; inflammatory bowel disease;

KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;  
 KW graft versus host disease; psoriasis; immunosuppressive; therapy;  
 KW complementarity determining region; CDR.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 24..34  
 FT /note= "CDR1"  
 FT Region 50..56  
 FT /note= "CDR2"  
 FT Region 89..97  
 FT /note= "CDR3"  
 XX  
 PN MO9743416-A1.  
 XX  
 PD 20-NOV-1997.  
 XX  
 PF 09-MAY-1997; 97MO-US07870.  
 XX  
 PR 10-MAY-1996; 96US-0017466.  
 XX  
 PA (BIOJ ) BIOGEN INC.  
 XX  
 PI Benjamin CD, Burkly LC, Hession C, Whitty A;  
 XX  
 DR WPI: 1998-008885/01.  
 DR N-PSDB; AAR97440.  
 XX  
 PT Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell anergy for  
 PT treatment of immunological diseases  
 XX  
 PS Claim 22: Page 81; 111pp; English.  
 XX  
 CC This polypeptide comprises the light chain variable region (VL) of  
 CC monoclonal antibody (MAb) C9.B8, which is produced by a hybridoma  
 CC deposited as ATCC 12107, and which is specific for the gamma  
 CC constant (gc) chain (see AAW31646) of human cytokine receptors. The  
 CC invention provides compositions and methods for inhibiting cytokine  
 CC signalling using gc chain blocking agents for the treatment of  
 CC immunological diseases such as myasthenia gravis, rheumatoid  
 CC arthritis, lupus, multiple sclerosis, insulin-dependent diabetes,  
 CC inflammatory bowel disease, sympathetic ophthalmia, uveitis,  
 CC allergy, asthma, parasitic infection, graft vs. host disease or  
 CC psoriasis. Preferred gc blocking agents include MAb CP.B8, its Fab  
 CC fragment and an antibody having a light chain variable region  
 CC CDR selected from those of CP.B8 VL or a heavy chain variable  
 CC region CDR selected from those of CP.B8 VH (see AAW31648).  
 XX  
 SQ Sequence 106 AA:  
 XX  
 QY 1 PWTFGGGTKL 10  
 DB 95 PWTFGGGTKL 104  
 XX  
 RESULT 5  
 ID AAR11985 standard; Protein; 107 AA.  
 XX  
 AC AAR11985;  
 XX  
 DT 31-JUL-1991 (first entry)  
 XX  
 DE Anti-placental alkaline phosphatase Vx.  
 XX  
 KW Placental alkaline phosphatase; antibody; cancer; light chain;  
 XX

KW PLAP; CDR.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 24..34  
 FT /label= CDR1  
 FT Region 50..56  
 FT /label= CDR2  
 FT Region 89..97  
 FT /label= CDR3  
 XX  
 PN EP429242-A.  
 XX  
 PD 29-MAY-1991.  
 XX  
 PF 14-NOV-1990; 90EP-0312407.  
 XX  
 PR 07-SEP-1990; 90GB-0019552.  
 PR 17-NOV-1989; 89GB-0026045.  
 XX  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Verhoeven ME;  
 XX  
 DR WPI: 1991-157662/22.  
 DR N-PSDB; AAO11846.  
 XX  
 PT Synthetic specific binder for human placental alkaline  
 PT phosphatase - for treating and diagnosing cancers e.g. breast,  
 PT ovarian and colon cancers and pleural effusions  
 XX  
 PS Disclosure; Fig 2; 28pp; English.  
 XX  
 CC The murine anti-PLAP Vx gene contains 3 CDRs which are used to reshape  
 CC the human anti-PLAP Vx gene (AAQ11848). The murine anti-PLAP Vx gene  
 CC (AAQ11845) CDRs were used to reshape the human anti-PLAP gene Vx  
 CC (AAQ11847). The specific binder for human PLAP contains at least  
 CC one of the former 3 CDRs and/or at least one of the latter CDRs.  
 CC The produced antibodies are more readily tolerated when administered  
 CC to a human patient. Antibody reagents can be used to identify, e.g.  
 CC by serum testing or imaging, and/or to treat PLAP-producing cancers.  
 CC Such cancers can occur as, e.g. breast cancer, ovarian cancer and  
 CC colon cancer or can manifest themselves as liquids such as pleural  
 CC effusions.  
 XX  
 SQ Sequence 107 AA:  
 XX  
 QY 1 PWTFGGGTKL 10  
 DB 95 PWTFGGGTKL 104  
 XX  
 RESULT 6  
 ID AAR25411 standard; Protein; 107 AA.  
 XX  
 AC AAR25411;  
 XX  
 DT 11-JAN-1993 (first entry)  
 XX  
 DE Light chain variable domain of MAb CB6 to TNFalpha epitope.  
 XX  
 KW monoclonal antibody; humanised; chimeric; CDR; neoplastic therapy;  
 KW shock; antilymphocyte therapy; endotoxin shock; septic shock;  
 KW sepsis; cardiovascular shock; tumour necrosis factor alpha;  
 KW multi-organ failure.  
 XX

```

CS Chimeric.
XX
XX WO9211383-A.
XX
XX 09-JUL-1992.
XX
XX 20-DEC-1991; 91WO-GB02300.
XX
XX 21-DEC-1990; 90WO-GB02017.
XX
XX 03-MAY-1991; 91GB-0009645.
XX
XX (CLLT ) CELLTech LTD.
XX
XX Adair JR, Althwal DS, Bodmer MW, Entage JS;
XX
XX WPI: 1992-250102/30.
XX
XX Recombinant antibody specific for human TNF-alpha - for treatment
XX of shock and immuno:regulatory and inflammatory disorders, also
XX used in diagnosis
XX
XX Claim 2; Fig 3; 57pp; English.
XX
XX This sequence is that of the light chain variable domain of a chimeric
XX antibody to an epitope of human tumour necrosis factor. The CDR-grafting
XX of this light chain has been based on the REI light chain. A gene
XX was built which coded for amino acid residues 23, 24, 48, 49, 71,
XX 73, 75, 88 [GH341(8)] as murine residues. It was then co-expressed
XX with g1221, the minimum grafted light chain (CDRs only).
XX
XX Sequence 107 AA:

Query Match      100.0%; Score 61; DB 13; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10
   |||||
DB 95 pwtfgggtkl 104

RESULT 7
AAR38614
ID AAR38614 standard; peptide: 107 AA.
XX
XX AAR38614;
XX
XX 28-OCT-1993 (first entry)
XX
XX Low-risk modified (prop) light chain of H65 Mab.
XX
XX Antibody: variable domain; light; L; heavy; H; consensus;
XX affinity: antigen; immunogenicity: humanisation; framework;
XX monoclonal antibody; Mab; hkl; h3.
XX
XX Mus sp.
XX
XX WO9311794-A.
XX
XX 24-JUN-1993.
XX
XX 14-DEC-1992; 92WO-US10906.
XX
XX 13-DEC-1991; 91US-0808464.
XX
XX (XOMA ) XOMA CORP.
XX
XX Fishwild DM, Kohn FR, Little RG, Studnicka GM;
XX
XX WPI: 1993-213827/26.
XX
XX Antibodies prepn. used for treatment of auto-immune diseases - by

```

```

PT replacement of critical residues to reduce immunogenicity but
PT retain binding affinity, etc.
XX
XX Example 1; Fig 6A; 160pp; English.
XX
XX Modified antibody variable domains were prepd. by identifying
XX low-risk residues in a mouse monoclonal antibody variable domain,
XX designated H65, which may be modified without diminishing the
XX native affinity of the domain for antigen while still reducing
XX its immunogenicity w.r.t. humans. The light and heavy chains of
XX the variable domain of H65 were determined to most closely resemble
XX the consensus sequences of subgroup 1 (hkl) of the human kappa
XX chains and subgroup 3 (h3) of the human heavy chains, respectively.
XX The H65 V/J-segments of the light and heavy chain sequences are given
XX in AAR38614 and AAR38615 respectively.
XX
XX Sequence 107 AA:

Query Match      100.0%; Score 61; DB 14; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10
   |||||
DB 95 pwtfgggtkl 104

RESULT 8
AAR41283
ID AAR41283 standard; peptide: 107 AA.
XX
XX AAR41283;
XX
XX 28-OCT-1993 (first entry)
XX
XX Moderate-risk modified (prop) light chain of H65 Mab.
XX
XX Antibody: variable domain; light; L; heavy; H; consensus;
XX affinity: antigen; immunogenicity: humanisation; framework;
XX monoclonal antibody; Mab; hkl; h3.
XX
XX Mus sp.
XX
XX WO9311794-A.
XX
XX 24-JUN-1993.
XX
XX 14-DEC-1992; 92WO-US10906.
XX
XX 13-DEC-1991; 91US-0808464.
XX
XX (XOMA ) XOMA CORP.
XX
XX Fishwild DM, Kohn FR, Little RG, Studnicka GM;
XX
XX WPI: 1993-213827/26.
XX
XX Antibodies prepn. used for treatment of auto-immune diseases - by
XX replacement of critical residues to reduce immunogenicity but
XX retain binding affinity, etc.
XX
XX Example 6; Fig 16A; 160pp; English.
XX
XX Modified antibody variable domains were prepd. by identifying
XX moderate-risk residues in a mouse monoclonal antibody variable domain,
XX designated H65, which may be modified without diminishing the
XX native affinity of the domain for antigen while still reducing
XX its immunogenicity w.r.t. humans. The light and heavy chains of
XX the variable domain of H65 were determined to most closely resemble
XX the consensus sequences of subgroup 1 (hkl) of the human kappa
XX chains and subgroup 3 (h3) of the human heavy chains, respectively.
XX The H65 V/J-segments of the light and heavy chain sequences are given

```



CC in AAR42183 and AAR42184 respectively.  
 XX  
 SQ Sequence 107 AA;

Query Match 100.0%; Score 61; DB 14; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTXL 10  
 |||||  
 Db 95 pwtfgggtkl 104

RESULT 9  
 AAM66098  
 ID AAM66098 standard; Protein; 107 AA.  
 XX  
 AC AAM66098;

DT 10-DEC-1998 (first entry)  
 XX

DE anti-CD22 monoclonal antibody light chain variable region.

XX anti-CD22 monoclonal antibody light chain variable region; VL;  
 KW Pseudomonas exotoxin; variable heavy chain; VH; variable light chain;  
 KM malignant B-cell; immunodiagnosis; RFB4 IgG.  
 XX

OS Mammalia.

XX  
 PN WO9841641-A1.

PD 24-SEP-1998.

PF 19-MAR-1998; 98WO-US05453;

PR 20-MAR-1997; 97US-0041437.

XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Fitzgerald D, Kreitman R, Mansfield E, Pastan I;

DR WPI; 1998-521227/44.

XX N-PSDB; AAV07641.

PT Recombinant anti-CD22 antibodies and immuno-conjugates - of  
 PT antibodies linked to a therapeutic agent, e.g. Pseudomonas exotoxin  
 PT or a label; for inhibiting malignant B-cells

XX  
 PS Claim 6; Fig 1; 71pp; English.

XX The invention claims for a recombinant immunconjugate comprising  
 CC of a therapeutic agent (e.g. Pseudomonas exotoxin) or a detectable  
 CC label peptide bonded to a recombinant anti-CD22 antibody (RFB4 IgG)  
 CC having a variable heavy (VH; AAM66099) chain with a cysteine residue  
 CC at amino acid 44 and the present variable light (VL) chain with a  
 CC cysteine residue at amino acid 100. The immunconjugate is claimed  
 CC to inhibit the growth of malignant B-cells in vivo, such as rodent,  
 CC canine or primate B-cells. The anti-CD22 antibody is claimed useful  
 CC for detecting CD22 protein in a sample or in vivo in a mammal, and  
 CC can be used in diagnostic kits.

XX  
 SQ Sequence 107 AA;

Query Match 100.0%; Score 61; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTXL 10  
 |||||  
 Db 95 pwtfgggtkl 104

RESULT 10  
 AAM49812  
 ID AAM49812 standard; Protein; 107 AA.  
 XX  
 AC AAM49812;

DT 24-SEP-1998 (first entry)  
 XX

DE Amino acid sequence of the mouse antibody C4G1 mature light chain.

XX Light chain; humanised; immunoglobulin; Ig; mouse C4G1; antibody;  
 KW inhibition; antigen; cardiovascular disease; thromboembolic disorder;  
 KW cancer; acute myocardial infarction; unstable angina; stroke;  
 KW transient ischemic episode; pulmonary embolism; deep vein thrombosis;  
 KM extracorporeal cardiopulmonary circulation.  
 XX

OS Mus sp.

XX  
 FH Key Location/Qualifiers

FT Domain 24..34 /note= "complementarity determining region"

FT Domain 50..56 /note= "complementarity determining region"

FT Domain 89..97 /note= "complementarity determining region"

XX US5777085-A.

XX 07-JUL-1998.

PF 17-MAY-1995; 95US-0458516.

PR 03-MAY-1993; 93US-0059159.

PR 20-DEC-1991; 91US-0812111.

PR 09-JUN-1992; 92US-0895952.

PR 11-SEP-1992; 92US-0944159.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Co MS, Tso JY;

DR WPI; 1998-398136/34.

XX New humanised immunoglobulin which binds GPIIb/IIIa - derived from  
 PT mouse C4G1 antibody, used for inhibiting platelet aggregation for  
 PT treating cardiovascular and thromboembolic disorders.

XX  
 PS Claim 1; Fig 5A; 35pp; English.

XX This is the amino acid sequence of the humanised antibody C4G1 light  
 CC chain, used in the method of the invention involving the creation  
 CC of a humanised immunoglobulin (Ig) derived from the mouse C4G1 antibody.  
 CC The humanised Ig is capable of binding to GPIIb/IIIa and inhibiting  
 CC platelet aggregation and also the releasing reaction of platelets. The  
 CC Ig can be used for treating cardiovascular diseases and thromboembolic  
 CC disorders, e.g. acute myocardial infarction, unstable angina, stroke,  
 CC transient ischemic episode, deep vein thrombosis and pulmonary embolism,  
 CC extracorporeal cardiopulmonary circulation. The Ig can also be used in  
 CC diagnosing the presence and location of a thrombus, or certain types of  
 CC cancer cells which develop GPIIb/IIIa on their surfaces, for the  
 CC detection of GPIIb/IIIa antigens or for isolating platelets.

XX  
 SQ Sequence 107 AA;

Query Match 100.0%; Score 61; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTXL 10  
 |||||  
 Db 95 pwtfgggtkl 104

```

RESULT 11
AAW58515
ID AAW58515 standard; Protein: 107 AA.
XX
XX AAW58515;
AC
XX
XX 18-AUG-1998 (first entry)
DT
XX
XX Protein SEQ ID NO:87 from US5770196.
DE
XX
XX Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;
KW depletion; cytotoxic; immunconjugate; fusion protein; psoriasis;
KW autoimmune disease; rheumatoid arthritis; type I diabetes.
XX
XX Unidentified.
OS
XX
XX US5770196-A.
PN
XX
XX 23-JUN-1998.
PD
XX
XX 07-JUN-1995; 95US-0472788.
PF
XX
XX 23-JUN-1993; 93US-0082842.
PR 13-DEC-1991; 91US-0808464.
PR 14-DEC-1992; 92WO-US10906.
PR 07-JUN-1995; 95US-0472788.
XX
XX (XOMA ) XOMA CORP.
PA
XX
XX Studnicka GM;
PI
XX
XX WPI: 1998-376744/32.
DR
XX
XX Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies
PT with humanised variable regions
PS
XX
XX Disclosure: Column 93-94; 77pp: English.
PS
XX
XX A method has been developed of depleting CD5+ cells in an animal. The
CC method comprises administering a cytotoxic protein containing a modified
CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig
CC molecule or an immunconjugate or fusion protein containing an anti-CD5
CC Ig molecule, and where the modified Ig variable domain comprises at
CC least one of (a) a modified light chain variable region (see AAW58478 or
CC AAW58480), and (b) a modified heavy chain variable region (see AAW58479
CC or AAW58481), where AAW58478 and AAW58479 are humanised forms of the H65
CC light and heavy chain variable domains with low risk amino acid
CC substitutions [i.e. low risk of reducing antigen-binding specificity.]
CC and AAW58480 and AAW58481 are humanised forms of the H65 light and heavy
CC chain variable domains with moderate risk amino acid substitutions and
CC are present in humanised H65 antibody he3 (ATCC HB 11206). The method is
CC useful for treating autoimmune diseases, especially systemic lupus
CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The
CC present sequence is given in the sequence listing of the present
CC invention but is not mentioned further within the specification.
CC
XX
XX Sequence 107 AA;
SQ

```

Query Match 100.0%; Score 61; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 PWTGGGTRL 10
   |||
Db 95 pwtl99gckl 104

```

RESULT 12  
 AAW58505  
 ID AAW58505 standard; Protein: 107 AA.

```

XX
AC AAW58505;
XX
XX 18-AUG-1998 (first entry)
DT
XX
XX H65 protein sequence SEQ ID NO:26 from Fig 6.
DE
XX
XX Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;
KW depletion; cytotoxic; immunconjugate; fusion protein; psoriasis;
KW autoimmune disease; rheumatoid arthritis; type I diabetes.
XX
XX Mus sp.
OS
XX
XX US5770196-A.
PN
XX
XX 23-JUN-1998.
PD
XX
XX 07-JUN-1995; 95US-0472788.
PF
XX
XX 23-JUN-1993; 93US-0082842.
PR 13-DEC-1991; 91US-0808464.
PR 14-DEC-1992; 92WO-US10906.
PR 07-JUN-1995; 95US-0472788.
XX
XX (XOMA ) XOMA CORP.
PA
XX
XX Studnicka GM;
PI
XX
XX WPI: 1998-376744/32.
DR
XX
XX Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies
PT with humanised variable regions
PS
XX
XX Example 6; Column 61-64; 77pp: English.
PS
XX
XX A method has been developed of depleting CD5+ cells in an animal. The
CC method comprises administering a cytotoxic protein containing a modified
CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig
CC molecule or an immunconjugate or fusion protein containing an anti-CD5
CC Ig molecule, and where the modified Ig variable domain comprises at
CC least one of (a) a modified light chain variable region (see AAW58478 or
CC AAW58480), and (b) a modified heavy chain variable region (see AAW58479
CC or AAW58481), where AAW58478 and AAW58479 are humanised forms of the H65
CC light and heavy chain variable domains with low risk amino acid
CC substitutions [i.e. low risk of reducing antigen-binding specificity.]
CC and AAW58480 and AAW58481 are humanised forms of the H65 light and heavy
CC chain variable domains with moderate risk amino acid substitutions and
CC are present in humanised H65 antibody he3 (ATCC HB 11206). The method is
CC useful for treating autoimmune diseases, especially systemic lupus
CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The
CC present sequence represents an H65 sequence from the present invention.
CC
XX
XX Sequence 107 AA;
SQ

```

Query Match 100.0%; Score 61; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 PWTGGGTRL 10
   |||
Db 95 pwtl99gckl 104

```

RESULT 13  
 AAW58478  
 ID AAW58478 standard; Protein: 107 AA.  
 AC AAW58478;  
 XX  
 XX 18-AUG-1998 (first entry)  
 DT  
 XX Humanised anti-CD5 immunoglobulin H65 light chain variable region #1.  
 DE

XX Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;  
 KW depletion; cytotoxic; immunconjugate; fusion protein; psoriasis;  
 KM autoimmune disease; rheumatoid arthritis; type I diabetes.  
 XX  
 OS Synthetic.  
 OS Mus sp.  
 OS Homo sapiens.  
 PN US5770196-A.  
 XX  
 PD 23-JUN-1998.  
 XX  
 PF 07-JUN-1995; 95US-0472788.  
 XX  
 PR 23-JUN-1993; 93US-0082842.  
 PR 13-DEC-1991; 91US-0808464.  
 PR 14-DEC-1992; 92WO-0510906.  
 PR 07-JUN-1995; 95US-0472788.  
 XX  
 PA (XOMA ) XOMA CORP.  
 XX  
 PI Studnicka GM;  
 XX  
 DR WPI; 1998-376744/32.  
 XX  
 PT Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies  
 PT with humanised variable regions  
 XX  
 PS Claim 1; Column 63-64; 77pp; English.  
 XX  
 CC A method has been developed of depleting CD5+ cells in an animal. The  
 CC method comprises administering a cytotoxic protein containing a modified  
 CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig  
 CC molecule or an immunconjugate or fusion protein containing an anti-CD5  
 CC Ig molecule, and where the modified Ig variable domain comprises at  
 CC least one of (a) a modified light chain variable region (see AAM58478 or  
 CC AAM58480), and (b) a modified heavy chain variable region (see AAM58479  
 CC or AAM58481), where AAM58478 and AAM58479 are humanised forms of the H65  
 CC light and heavy chain variable domains with low risk amino acid  
 CC substitutions (i.e. low risk of reducing antigen-binding specificity.)  
 CC and AAM58480 and AAM58481 are humanised forms of the H65 light and heavy  
 CC chain variable domains with moderate risk amino acid substitutions and  
 CC are present in humanised H65 antibody hex (ATCC HB 11206). The method is  
 CC useful for treating autoimmune diseases, especially systemic lupus  
 CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The  
 CC present sequence represents a specifically claimed humanised anti-CD5  
 CC immunoglobulin H65 light chain variable region.  
 CC  
 SQ Sequence 107 AA;  
 OY  
 Query Match 100.0%; Score 61; DB 19; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
 1 PMTFGGGKTL 10  
 |||||||||  
 Db 95 pwtfgggtkl 104  
 RESULT 14  
 AAR21292  
 ID AAR21292 standard; Protein; 108 AA.  
 XX  
 AC .AAR21292;  
 XX  
 DT 21-MAY-1992 (first entry)  
 XX  
 DE Murine VL kappa group VI chain "g", specific for phox.  
 XX  
 KW Rd; bacteriophage; gene III; filamentous; phagemid; capsid; coat;  
 KW plus; g3p; binding; adsorption; gene VIII; diverse repertoire;

KW specific binding pairs; replicable genetic display package.  
 XX  
 OS Synthetic.  
 XX  
 FH Key  
 FT Binding-site 24..33 Location/Qualifiers  
 FT Binding-site /label= CDR1  
 FT Binding-site 49..55  
 FT Binding-site /label= CDR2  
 FT Binding-site 88..96  
 FT /label= CDR3  
 FT /note= "D-X-G-X-X motif"  
 PN WO9201047-A.  
 XX  
 PD 23-JAN-1992.  
 XX  
 PF 10-JUL-1991; 91WO-GB01134.  
 XX  
 PR 15-MAY-1991; 91GB-0010549.  
 PR 10-JUL-1990; 90GB-0015198.  
 PR 19-OCT-1990; 90GB-0022845.  
 PR 12-NOV-1990; 90GB-0024503.  
 PR 06-MAR-1991; 91GB-0004744.  
 XX  
 PA (CMB-) CAMBRIDGE ANTIBODY.  
 PA (MED-) MED RES COUNCIL.  
 XX  
 PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;  
 PI Jackson RH, Holliger KP, Marks JD, Clackson TP, Chiswell DJ;  
 PI Winter GP, Bonnett TP;  
 XX  
 DR WPI; 1992-056862/07.  
 XX  
 PT Producing members of specific binding pairs - by expression in  
 PT recombinant host cells with a secreting replicable genetic  
 PT display package.  
 XX  
 PS Example 21; Fig 24; 209pp; English.  
 XX  
 CC The VK sequence is one of seven (AAR21286-92) found to be expressed  
 CC from a single chain Fv library from an immunised mouse. The libra-  
 CC ry produces a diverse repertoire of antibody fragments specific for  
 CC 2-phenyl-5-oxazolone (phox). It was prepd. using cDNA generated  
 CC from mRNA from mice immunised with phox coupled to chicked serum  
 CC albumin. The VH and VL kappa sequences were separately amplified  
 CC by PCR (see AAQ23474-84) and ligated into fdCMT2 (see AAQ23463) for ex-  
 CC pression on the phage surface as fusions with gene III. The result-  
 CC ing library of clones was diverse. Twenty three hapten binding  
 CC clones were sequenced revealing eight different VH genes (A-H) (see  
 CC AAR21264-71) in a variety of pairings with the seven different VK  
 CC genes (a-g). Of the twenty three clones sequenced, only one was of  
 CC type "g", and was an "ox-like" gene. (See Berex et al, Nature 316  
 CC 412-418, 1985). It contains the DXGXX motif in CDR3, the central  
 CC Gly of which is needed to create a cavity for phox. Most of the  
 CC clones were VK-d combinations. The Kd of VH-B/VK-b for phox-GABA  
 CC was 10 uM, one of the highest values found. This suggests that  
 CC phage bearing scFv fragments having weak affinities can be selected  
 CC with antigen, probably due to the avidity of the multiple antibody  
 CC heads on the phage.  
 CC See also AAR21260-307, 309-311; AAR22450, 565-581.  
 CC  
 SQ Sequence 108 AA;  
 OY  
 Query Match 100.0%; Score 61; DB 13; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
 1 PMTFGGGKTL 10  
 |||||||||  
 Db 94 pwtfgggtkl 103

RESULT 15  
 AAR21310  
 ID AAR21310 standard; Protein: 108 AA.  
 AC AAR21310;  
 XX  
 DT 21-MAY-1992 (first entry)  
 XX  
 DE Light chain of M1f clone.  
 XX  
 KW Fd: bacteriophage; gene III; filamentous; phagemid; capsid; coat;  
 KW plus: g3p; binding; adsorption; gene VIII; diverse repertoire;  
 KW specific binding pairs; replicable genetic display package.  
 XX  
 OS Synthetic.  
 XX  
 FH Key  
 FT Region 25..34 Location/Qualifiers  
 FT /label= CDR1  
 FT 50..56  
 FT /label= CDR2  
 FT 89..96  
 FT Region /label= CDR3  
 FT  
 XX  
 XX W09201047-A.  
 XX  
 XX 23-JAN-1992.  
 XX  
 PD  
 XX  
 PD 10-JUL-1991; 91WO-GB01134.  
 XX  
 PF 15-MAY-1991; 91GB-0010549.  
 XX  
 PR 10-JUL-1990; 90GB-0015198.  
 PR 19-OCT-1990; 90GB-0022845.  
 PR 12-NOV-1990; 90GB-0024503.  
 PR 06-MAR-1991; 91GB-0004744.  
 XX  
 PA (CMB-) CAMBRIDGE ANTIBODY.  
 PA (MED-) MED RES COUNCIL.  
 XX  
 PI McCafferty J, Pope AR, Johnson KS, Hoogenboom HRJ, Griffiths AD;  
 PI Jackson RH, Holliger KP, Marks JD, Jackson TP, Chiswell DJ;  
 PI Winter GP, Bonnett TP;  
 XX  
 DR WPI: 1992-056862/07.  
 XX  
 XX Producing members of specific binding pairs - by expression in  
 PT recombinant host cells with a secreting replicable genetic  
 PT display package.  
 XX  
 PS Example 46; Fig 52; 109pp; English.  
 XX  
 CC The sequence is the light chain of clone M1f encoding an scFv frag-  
 CC ment specific for both hen and turkey egg lysozyme (HEL and TEL).  
 CC The DNA encoding the chain was amplified from a cDNA library prepd.  
 CC from the spleen of an unimmunised mouse. The corresponding heavy  
 CC chain was also amplified from an existing construct, PSW1-VHD1.3  
 CC (Ward et al, 1989). The two fragments were assembled via a linker  
 CC vector for expression on the surface of fd bacteriophage. In this  
 CC way, the VL domain was replaced by a library of VL domains to allow  
 CC for selection of a broader range of antibody specificities. Several  
 CC clones were isolated which bound to TEL (the parent antibody D1.3  
 CC binds exclusively to HEL). The sequences of the light chains of  
 CC two of these clones, M1 and M21 are given in AAR21310 and AAR21311  
 CC respectively. The D1.3 light chain is given in AAR21309.  
 CC See also AAR21260-307, 308-312, AAR22450, AAR22565, AAR22567-81.  
 XX  
 SQ Sequence 108 AA;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PWTFGGTRK 10  
 DB 95 pwtfggtrkl 104

Search completed: August 20, 2002, 13:15:11  
 Job time: 105 sec

Query Match 100.0%; Score 61; DB 13; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 0.035;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: August 20, 2002, 13:16:03 ; Search time 46.28 Seconds  
(without alignments)  
20.763 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_95\_104  
Perfect score: 61  
Sequence: 1 PMTFGGGCTKL 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	77	2 G28840	Ig kappa chain V r
2	61	100.0	79	2 A28840	Ig kappa chain V r
3	61	100.0	88	2 PL0261	Ig kappa chain V r
4	61	100.0	95	2 S26333	Ig kappa chain V r
5	61	100.0	96	2 G38601	Ig kappa chain V r
6	61	100.0	98	2 S19974	Ig kappa chain V r
7	61	100.0	103	2 S18731	Ig kappa chain V-J
8	61	100.0	106	2 PL0259	Ig kappa chain V r
9	61	100.0	106	2 PL0259	Ig kappa chain V r
10	61	100.0	106	2 PL0262	Ig kappa chain V r
11	61	100.0	107	2 S09964	Ig kappa chain V-J
12	61	100.0	107	2 PL0268	Ig kappa chain V r
13	61	100.0	107	2 S09968	Ig kappa chain V-J
14	61	100.0	108	2 S19112	Ig kappa chain V r
15	61	100.0	108	2 S38862	Ig kappa chain V r
16	61	100.0	108	2 S69900	Ig kappa chain V r
17	61	100.0	109	2 PH0093	Ig kappa chain V r
18	61	100.0	109	2 S26336	Ig kappa chain V r
19	61	100.0	110	2 PH0090	Ig light chain V r
20	61	100.0	111	1 KVM880	Ig kappa chain V r
21	61	100.0	111	2 E38740	Ig kappa chain V r
22	61	100.0	111	2 A38740	Ig kappa chain V r
23	61	100.0	111	2 G38740	Ig kappa chain V r
24	61	100.0	111	2 C38740	Ig kappa chain V r
25	61	100.0	112	2 A27887	Ig kappa chain V r
26	61	100.0	112	2 C27887	Ig kappa chain V r
27	61	100.0	112	2 A49715	Ig kappa chain V r
28	61	100.0	112	2 B31485	Ig kappa chain V r
29	61	100.0	112	2 A36259	Ig kappa chain V r

30	61	100.0	112	2 PL0273	Ig kappa chain V r
31	61	100.0	112	2 S38716	Ig light chain V r
32	61	100.0	113	2 E30560	Ig kappa chain V r
33	61	100.0	114	2 PT0358	Ig kappa chain V r
34	61	100.0	118	2 PT0359	Ig kappa chain V r
35	61	100.0	121	2 S67944	Ig kappa chain B r
36	61	100.0	122	2 A29380	Ig kappa chain pre
37	61	100.0	126	2 A34904	Ig kappa chain pre
38	61	100.0	127	2 S04577	Ig kappa chain pre
39	61	100.0	127	2 S04574	Ig kappa chain pre
40	61	100.0	128	2 A47159	Ig lambda chain V
41	61	100.0	130	1 KVM5M4	Ig kappa chain pre
42	61	100.0	131	1 KVM5M6	Ig kappa chain pre
43	61	100.0	131	2 B34904	Ig kappa chain pre
44	61	100.0	131	2 C34904	Ig kappa chain pre
45	61	100.0	131	2 S52449	Ig kappa chain V r

## ALIGNMENTS

RESULT 1  
G28840  
Ig kappa chain V region (HP25) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 23-Jul-1999  
C:Accession: G28840: 025114  
R:Ollier, P.; Rocca-Serra, J.; Somme, G.; Theze, J.; Fougereau, M.  
EMBO J. 4, 3681-3688, 1985  
A:Title: The idiotypic network and the internal image: possible regulation of a germ-  
A:Reference number: A91028; MUID:86136012  
A:Accession: G28840  
A:Molecule type: mRNA  
A:Residues: 1-77 <OLN>  
A:Cross-references: GB:X03389; NID:952174; PIDN:CAA27120.1; PID:952175  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 61; DB 2; Length 77;  
Best Local Similarity 100.0%; Pred. No. 0.0028;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PMTFGGGCTKL 10  
DB 65 PMTFGGGCTKL 74

RESULT 2  
A28840  
Ig kappa chain V region (HP21) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 07-Mar-1988 #sequence\_revision 07-Mar-1988 #text\_change 23-Jul-1999  
C:Accession: A28840: M25114  
R:Ollier, P.; Rocca-Serra, J.; Somme, G.; Theze, J.; Fougereau, M.  
EMBO J. 4, 3681-3688, 1985  
A:Title: The idiotypic network and the internal image: possible regulation of a germ-  
A:Reference number: A91028; MUID:86136012  
A:Accession: A28840  
A:Molecule type: mRNA  
A:Residues: 1-79 <OLN>  
A:Cross-references: GB:X03387; NID:952168; PIDN:CAA27118.1; PID:952169  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotetramer; Immunoglobulin

Query Match 100.0%; Score 61; DB 2; Length 79;  
Best Local Similarity 100.0%; Pred. No. 0.0029;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PMTFGGGCTKL 10  
DB 67 PMTFGGGCTKL 76

RESULT 3  
PL0261  
Ig kappa chain V region (anti-DNA, DP15VK) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 09-May-1997  
C:Accession: PL0261  
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.  
J. Exp. Med. 171, 265-297, 1990  
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic  
A:Reference number: PL0231; MUID:90111618  
A:Accession: PL0261  
A:Molecule type: mRNA  
A:Residues: 1-88 <SHL>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-5/Region: framework 1  
F:6-16/Region: complementarity-determining 1  
F:17-31/Region: framework 2  
F:32-38/Region: complementarity-determining 2  
F:39-70/Region: framework 3  
F:71-79/Region: complementarity-determining 3  
F:80-88/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 88;  
Best Local Similarity 100.0%; Pred. No. 0.0032;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10  
|||||  
DB 77 PWTGGGTRKL 86

RESULT 4  
S26333  
Ig kappa chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Jun-2000  
C:Accession: S26333  
R:Stark, S.E.; Caton, A.J.  
J. Exp. Med. 174, 613-624, 1991  
A:Title: Antibodies that are specific for a single amino acid interchange in a protein  
A:Reference number: S26309; MUID:91341421  
A:Accession: S26333  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-95 <STRA>  
A:Cross-references: EMBL:X59181; NID:9552312; PID:91334061  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:6-80/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 0.0034;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10  
|||||  
DB 85 PWTGGGTRKL 94

RESULT 5  
G38601  
Ig kappa chain V region (4A9) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 30-Aug-1991 #sequence\_revision 30-Aug-1991 #text\_change 23-Jul-1999  
C:Accession: G38601  
R:Goshorn, S.C.; Retzel, E.; Jemerson, R.  
J. Biol. Chem. 266, 2134-2142, 1991  
A:Title: Common structural features among monoclonal antibodies binding the same antigen

A:Reference number: A38601; MUID:91115823  
A:Accession: G38601  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-96 <GOS>  
A:Cross-references: GB:M57984; NID:9196414; PID:AAA63365.1; PID:9196415  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 61; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 0.0035;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10  
|||||  
DB 83 PWTGGGTRKL 92

RESULT 6  
S19974  
Ig kappa chain V region (M-T406) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 21-Jan-2000  
C:Accession: S19974  
R:Weissenhorn, W.; Rietmuller, G.; Weiss, E.M.; Rieber, E.P.  
Submitted to the EMBL Data Library, March 1992  
A:Description: Structural characterization of CD4 mab.  
A:Reference number: S19963  
A:Accession: S19974  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-98 <WEI>  
A:Cross-references: EMBL:X65096; NID:952294; PID:CAA6224.1; PID:952295  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-80/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 98;  
Best Local Similarity 100.0%; Pred. No. 0.0035;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRKL 10  
|||||  
DB 85 PWTGGGTRKL 94

RESULT 7  
S18731  
Ig kappa chain V-J region (MSI-N17) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 19-Mar-1997 #sequence\_revision 05-Dec-1998 #text\_change 23-Jul-1999  
C:Accession: S18731  
R:Hirama, T.; Takeshita, S.; Yoshida, Y.; Yamaagishi, H.  
Immunol. Lett. 27, 19-24, 1991  
A:Title: Structure of extrachromosomal circular DNAs generated by immunoglobulin light  
A:Reference number: S18731; MUID:91209891  
A:Accession: S18731  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-103 <HIR>  
A:Cross-references: EMBL:X54753; NID:955294; PID:CAA38555.1; PID:955295  
A:Experimental source: spleen, strain BALB/c-nu/nu  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1990  
C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (hain disulfide bonds. In some cases, such as Iga and Igm, the subunits associate into  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:11-85/Domain: immunoglobulin homology <IMM>  
F:18-83/Disulfide bonds: #status predicted

Query Match 100.0%; Score 61; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 0.0037;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||||  
DB 90 PWTGGGCTKL 99

## RESULT 8

PL0260  
Ig kappa chain V region (anti-DNA, DP7VK) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: PL0260  
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.  
J. Exp. Med. 171, 265-297, 1990  
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic  
A:Reference number: PL0231; MVID:90111618  
A:Accession: PL0260  
A:Molecule type: mRNA  
A:Residues: 1-106 <SHL>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:1-23/Region: framework 1  
F:16-90/Domain: immunoglobulin homology <IMM>  
F:24-34/Region: complementarity-determining 1  
F:35-49/Region: framework 2  
F:50-56/Region: complementarity-determining 2  
F:57-88/Region: framework 3  
F:89-97/Region: complementarity-determining 3  
F:98-106/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||||  
DB 95 PWTGGGCTKL 104

## RESULT 9

PL0259  
Ig kappa chain V region (anti-DNA, DP11VK) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: PL0259  
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.  
J. Exp. Med. 171, 265-297, 1990  
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic  
A:Reference number: PL0231; MVID:90111618  
A:Accession: PL0259  
A:Molecule type: mRNA  
A:Residues: 1-106 <SHL>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:1-23/Region: framework 1  
F:16-90/Domain: immunoglobulin homology <IMM>  
F:24-34/Region: complementarity-determining 1  
F:35-49/Region: framework 2  
F:50-56/Region: complementarity-determining 2  
F:57-88/Region: framework 3  
F:89-97/Region: complementarity-determining 3  
F:98-106/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10

DB 95 PWTGGGCTKL 104  
|||||

## RESULT 10

PL0262  
Ig kappa chain V region (anti-DNA, 60VK) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: PL0262  
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein, A.  
J. Exp. Med. 171, 265-297, 1990  
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic  
A:Reference number: PL0231; MVID:90111618  
A:Accession: PL0262  
A:Molecule type: mRNA  
A:Residues: 1-106 <SHL>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:1-23/Region: framework 1  
F:16-90/Domain: immunoglobulin homology <IMM>  
F:24-34/Region: complementarity-determining 1  
F:35-49/Region: framework 2  
F:50-56/Region: complementarity-determining 2  
F:57-88/Region: framework 3  
F:89-97/Region: complementarity-determining 3  
F:98-106/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||||  
DB 95 PWTGGGCTKL 104

## RESULT 11

S09964  
Ig kappa chain V-J region (105-2H) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 21-Jan-2000  
C:Accession: S09964  
R:Reininger, L.; Shibata, T.; Ozaki, S.; Shirai, T.; Jaton, J.C.; Izui, S.  
Eur. J. Immunol. 20, 771-777, 1990  
A:Title: Variable region sequences of pathogenic anti-mouse red blood cell autoantibodies  
A:Reference number: S09955; MVID:90269328  
A:Accession: S09964  
A:Molecule type: mRNA  
A:Residues: 1-107 <REI>  
A:Cross-references: EMBL:X51852; NID:g55393; PIDN:CA36145.1; PID:g930229  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin  
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.0039;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||||  
DB 95 PWTGGGCTKL 104

## RESULT 12

PL0268  
Ig kappa chain V region (anti-DNA, DP9VK and DP17VK) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jan-2000  
C:Accession: PL0268  
R:Shlomchik, M.; Mascelli, M.; Shan, H.; Radic, M.Z.; Pisetsky, D.; Marshak-Rothstein

J. Exp. Med. 171, 265-297, 1990  
A:Title: Anti-DNA antibodies from autoimmune mice arise by clonal expansion and somatic  
A:Reference number: PL0231; MUID:90111618  
A:Accession: PL0268  
A:Molecule type: mRNA  
A:Residues: 1-107 <SHL>  
C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:1-23/Region: framework 1  
F:16-90/Domain: immunoglobulin homology <IMM>  
F:24-34/Region: complementarity-determining 1  
F:35-49/Region: framework 2  
F:50-56/Region: complementarity-determining 2  
F:57-68/Region: framework 3  
F:89-97/Region: complementarity-determining 3  
F:98-107/Region: framework 4

Query Match 100.0%; Score 61; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.0039;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||  
Db 95 PWTGGGCTKL 104

RESULT 13  
S09968  
Ig kappa chain V-J region (34-2B) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 21-Jan-2000  
R:Accession: S09968  
R:Reininger, L.; Shihata, T.; Ozaki, S.; Shirai, T.; Jaton, J.C.; Izui, S.  
Eur. J. Immunol. 20, 771-777, 1990  
A:Title: Variable region sequences of pathogenic anti-mouse red blood cell autoantibodies  
A:Reference number: S09955; MUID:9026328  
A:Accession: S09968  
A:Molecule type: mRNA  
A:Residues: 1-107 <REI>  
A:Cross-references: EMBL:X51856; NID:955403; PIDN:CA36149.1; PID:9930233  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 0.0039;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||  
Db 95 PWTGGGCTKL 104

RESULT 14  
S19112  
Ig kappa chain V region - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jan-2000  
R:Accession: S19112  
R:Mylvaganam, S.E.; Paterson, Y.; Kaiser, K.; Bowdlish, K.; Getzoff, E.D.  
J. Mol. Biol. 221, 455-462, 1991  
A:Title: Biochemical implications from the variable gene sequences of an anti-cytochrome  
forms.  
A:Reference number: S17586; MUID:92015240  
A:Accession: S19112  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-108 <MTL>  
A:Cross-references: EMBL:X60684  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 108;  
Best Local Similarity 100.0%; Pred. No. 0.0039;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||  
Db 95 PWTGGGCTKL 104

RESULT 15  
S38862  
Ig kappa chain V region - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 21-Jan-2000  
R:Accession: S38862  
R:Fischer, R.; Voss, A.; Hunziker, W.; Stierhof, Y.D.; Kreuzaler, F.  
submitted to the EMBL Data Library, August 1993  
A:Description: Production and cloning of TMV-specific monoclonal antibodies.  
A:Reference number: S37200  
A:Accession: S38862  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-108 <FTS>  
A:Cross-references: EMBL:X75854; NID:9429109; PID:9429110  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin  
F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 100.0%; Score 61; DB 2; Length 108;  
Best Local Similarity 100.0%; Pred. No. 0.0039;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGCTKL 10  
|||  
Db 96 PWTGGGCTKL 105

Search completed: August 20, 2002, 13:16:04  
Job time: 158 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 13:17:52 ; Search time 21.48 seconds

(without alignments)  
18.026 Million cell updates/sec

Title: US-09-824-286-4\_COPY\_95\_104  
Perfect score: 61  
Sequence: 1 PWTFGGCTKL 10

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	111	1	KV3A_MOUSE
2	61	100.0	111	1	KV3C_MOUSE
3	61	100.0	130	1	KV3G_MOUSE
4	61	100.0	131	1	KV3I_MOUSE
5	55	90.2	110	1	KV3P_MOUSE
6	55	90.2	111	1	KV3R_MOUSE
7	55	90.2	111	1	KV3O_MOUSE
8	55	90.2	111	1	KV3Q_MOUSE
9	55	90.2	112	1	KV3G_MOUSE
10	55	90.2	132	1	KV3F_MOUSE
11	54	88.5	107	1	KV6A_MOUSE
12	52	85.2	108	1	KV5Q_MOUSE
13	52	85.2	108	1	KV5R_MOUSE
14	52	85.2	108	1	KV5S_MOUSE
15	52	85.2	108	1	KV5T_MOUSE
16	52	85.2	108	1	KV5U_MOUSE
17	52	85.2	111	1	KV3D_MOUSE
18	52	85.2	111	1	KV3H_MOUSE
19	52	85.2	111	1	KV3J_MOUSE
20	52	85.2	111	1	KV3L_MOUSE
21	52	85.2	112	1	KV2D_MOUSE
22	52	85.2	112	1	KV3B_MOUSE
23	52	85.2	113	1	KV2E_MOUSE
24	52	85.2	113	1	KV2F_MOUSE
25	52	85.2	113	1	KV2G_MOUSE
26	50	82.0	108	1	KV1B_HUMAN
27	50	82.0	108	1	KV1R_HUMAN
28	50	82.0	129	1	KV3M_HUMAN
29	50	82.0	134	1	KV4C_HUMAN
30	49	80.3	110	1	KV1C_HUMAN
31	49	80.3	111	1	KV6C_HUMAN
32	49	80.3	112	1	KV6B_HUMAN
33	49	80.3	129	1	KV1B_MOUSE

## ALIGNMENTS

RESULT ID	Query Match	Score	DB 1	Length	DB 111
1	100.0%	61	111	111	
2	100.0%	61	111	111	
3	100.0%	61	111	111	
4	100.0%	61	111	111	
5	100.0%	61	111	111	
6	100.0%	61	111	111	
7	100.0%	61	111	111	
8	100.0%	61	111	111	
9	100.0%	61	111	111	
10	100.0%	61	111	111	
11	100.0%	61	111	111	
12	100.0%	61	111	111	
13	100.0%	61	111	111	
14	100.0%	61	111	111	
15	100.0%	61	111	111	
16	100.0%	61	111	111	
17	100.0%	61	111	111	
18	100.0%	61	111	111	
19	100.0%	61	111	111	
20	100.0%	61	111	111	
21	100.0%	61	111	111	
22	100.0%	61	111	111	
23	100.0%	61	111	111	
24	100.0%	61	111	111	
25	100.0%	61	111	111	
26	100.0%	61	111	111	
27	100.0%	61	111	111	
28	100.0%	61	111	111	
29	100.0%	61	111	111	
30	100.0%	61	111	111	
31	100.0%	61	111	111	
32	100.0%	61	111	111	
33	100.0%	61	111	111	

Q	Mus musculus (Mouse).
O	Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mu-
CX	NCBI_TaxID=10090;
FN	[1]
RX	SEQUENCE.
FX	MEDLINE=67056897; PubMed=4162931;
RA	Gray W.R., Dreyer W.J., Hood L.E.;
RT	"Mechanism of antibody synthesis: size differences between mouse
PT	kappa chains.";
RL	Science 155:465-467(1967).
CC	-1- MISCELLANEOUS: THIS IS A BENCE-JONES PROTEIN.
DR	PIR: A01930; KMS80;
DR	HSP: P01789; IMCP.
DR	InterPro: IPR003006; Ig_MHC.
DR	InterPro: IPR003596; Ig_V.
DR	Pfam: PF00047; Ig_1.
KM	SMART: SM00406; IGV: 1.
FY	Immunoglobulin V region; Bence-Jones protein.
FT	DOMAIN 1 23 FRAMEWORK-1.
FT	DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT	DOMAIN 39 53 FRAMEWORK-2.
FT	DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT	DOMAIN 61 92 FRAMEWORK-3.
FT	DOMAIN 93 101 COMPLEMENTARITY-DETERMINING-3.
FT	DOMAIN 102 111 FRAMEWORK-4.
FT	DISULFID 23 - 92 BY SIMILARITY.
FT	NON_TER 111 111
SQ	SEQUENCE 111 AA; 11904 MW; 4EE7ABC9DF0FC125 CRC64;

```

Query Match      100.0%  Score 61;  DB 1;  Length 11;
Best Local Similarity 100.0%  Pred. No. 0.0011;
Matches 10;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0.

QY      1  PWTGGGCTKL 10
          |||||
D3      99  PWTGGGCTKL 108

```

RESULT	3	
KV5G_MOUSE		
ID	KV5G_MOUSE	STANDARD; PRT; 130 AA.
AC	P01639; P01640;	
DT	21-JUL-1986 (Rel. 01, Created)	
DT	21-JUL-1986 (Rel. 01, Last sequence update)	
DT	15-JUL-1999 (Rel. 38, Last annotation update)	
DE	ig kappa chain V-V region MOPC 41 precursor.	
OS	Mus musculus (Mouse)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
NCBI	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
NCBI	"taxid:10090";	
RN	[1]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=79221900; Pubmed=111146;	
RA	Seidman J.G., Max E.E., Leder P.;	
RT	"A kappa-immunoglobulin gene is formed by site-specific recombination	
RL	without further somatic mutation.";	
RL	Nature 280:370-375(1979).	
RN	[2]	
RN	SEQUENCE OF 1-33.	
RP	MEDLINE=77148916; Pubmed=403522;	
RX	Burstein Y., Schechter I.;	
RA	"Amino acid sequence of the NH2-terminal extra piece segments of the	
RT	precursors of mouse immunoglobulin lambdaI-type and kappa-type light	
RT	chains.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 74:716-720(1977).	
RL	[3]	
RP	SEQUENCE OF 23-130.	
RX	MEDLINE=67056897; Pubmed=4162931;	
RA	Gray W.R., Dreyer W.J., Hood L.;	
RT	"mechanism of antibody synthesis: size differences between mouse	
RT	kappa chains.";	

RL	science 155:465-467 (1967).
CC	-1- MISCELLANEOUS: THIS PRECURSOR WAS SYNTHESIZED IN A CELL-FREE
CC	SYSTEM DIRECTED BY MNA ISOLATED FROM MYELOMA POLYSOMES.
CC	-1- MISCELLANEOUS: THIS IS A BENGE-JONES PROTEIN.
CC	PtR; A01922; KWSM4.
DR	InterPro; IPR003006; Ig_MHC.
DR	InterPro; IPR003596; Ig_V.
DR	Pfam; PF00047; Ig; 1.
DR	SMART; SM00406; IGv; 1.
KW	Immunoglobulin V region; Signal; Bence-Jones protein.
FT	SIGNAL 1 22
FT	CHAIN 23 130
FT	D0MAIN 23 45
FT	D0MAIN 46 56
FT	D0MAIN 57 71
FT	D0MAIN 72 78
FT	D0MAIN 79 110
FT	D0MAIN 111 119
FT	D0MAIN 120 129
FT	DISULFID 45 110
FT	VARIANT 1 2
FT	NON_TER 130 130
SO	SEQUENCE 130 AA; 14311 MW; 5EEFE0FE7D5F1BEC CRC64;

```

Query Match          100.0%; Score 61; DB 1; length 130;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 PWTGGGTTKL 10
Db 117 PWTGGGTTKL 126

```

RESULT	4
AC	KV3L_MOUSE
ID	KV3L_MOUSE . STANDARD; PRT; 131 AA.
AC	P01611;
DT	21-JUL-1986 (Rel. 01, Created)
DT	21-JUL-1986 (Rel. 01, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	Ig kappa chain V-II1 region MOPC 63 precursor.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_Taxid=10090;
RN	[1]
RP	SEQUENCE OF 1-35.
RX	MEDLINE=7823587; PubMed=98179;
RA	Burstein Y., Schechter I.
RT	"Primary structures of N-terminal extra peptide segments linked to
RT	the variable and constant regions of immunoglobulin light chain
RT	precursors: implications on the organization and controlled
RT	expression of immunoglobulin genes.;"
RL	Biochemistry 17:2392-2400(1978).
RN	[2]
RP	SEQUENCE OF 21-131.
RX	MEDLINE=73140225; PubMed=4691517;
RA	McKean D.J., Potter M., Hood L.E.;
RT	"Mouse immunoglobulin chains. Pattern of sequence variation among
RT	kappa chains with limited sequence differences.;"
RL	Biochemistry 12:760-771(1973).
RN	[3]
RP	REVISIONS.
RX	MEDLINE=79012520; PubMed=99744;
RA	McKean D.J., Bell M., Potter M.;
RT	"Mechanisms of antibody diversity: multiple genes encode structurally
RT	related mouse kappa variable regions.;"
RL	Proc. Natl. Acad. Sci. U.S.A. 75:3913-3917(1978).
DR	PIR; A01935; KVM5M6.
DR	HSSP; P01789; 1MCP.
DR	InterPro; IPR003006; Ig_MHC.
DR	InterPro; IPR003596; Ig_V.

DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; Igv; 1.  
 KW Immunoglobulin V region; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 131 IG KAPPA CHAIN V-III REGION MOPC 63.  
 FT DOMAIN 21 43 FRAMEWORK-1.  
 FT DOMAIN 44 58 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 59 73 FRAMEWORK-2.  
 FT DOMAIN 74 80 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 81 112 FRAMEWORK-3.  
 FT DOMAIN 113 121 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 122 131 FRAMEWORK-4.  
 FT DISULFID 43 112 BY SIMILARITY.  
 FT NON\_TER 131 131  
 SQ SEQUENCE 131 AA; 14291 MW; D212EC9F08DC880A CRC64;

Query Match 100.0%; Score 61; DB 1; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 0.0013;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10  
 DB 119 PWTGCGTKL 128

RESULT 5  
 ID KV3P\_MOUSE STANDARD; PRT; 110 AA.  
 AC P01667;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-III region PC 7210.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79073152; PubMed=103003;  
 RA Weigert M., Gattalitan L., Loh E., Schilling J., Hood L.E.;  
 RT "Rearrangement of genetic information may produce immunoglobulin  
 RT diversity.";  
 RL Nature 276:785-790(1978).  
 RL PIR: D01937; KWS10.  
 DR HSSP; P01789; IMCP.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; Igv; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 39 53 FRAMEWORK-2.  
 FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 61 92 FRAMEWORK-3.  
 FT DOMAIN 93 100 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 101 110 FRAMEWORK-4.  
 FT DISULFID 23 92 BY SIMILARITY.  
 FT NON\_TER 110 110  
 SQ SEQUENCE 110 AA; 11950 MW; 69F1A5CE886B1249 CRC64;

Query Match 90.2%; Score 55; DB 1; Length 110;  
 Best Local Similarity 90.0%; Pred. No. 0.011;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10  
 DB 98 PWTGCGTKL 107

RESULT 6  
 ID KV3E\_MOUSE STANDARD; PRT; 111 AA.  
 AC P01657;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-III region PC 2413.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79073152; PubMed=103003;  
 RA Weigert M., Gattalitan L., Loh E., Schilling J., Hood L.E.;  
 RT "Rearrangement of genetic information may produce immunoglobulin  
 RT diversity.";  
 RL Nature 276:785-790(1978).  
 RL PIR: A01932; KWS13.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; Igv; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.  
 FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 39 53 FRAMEWORK-2.  
 FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 61 92 FRAMEWORK-3.  
 FT DOMAIN 93 101 COMPLEMENTARITY-DETERMINING-3.  
 FT DOMAIN 102 111 FRAMEWORK-4.  
 FT DISULFID 23 92 BY SIMILARITY.  
 FT NON\_TER 111 111  
 SQ SEQUENCE 111 AA; 11949 MW; ADB89EFC6EC746A9 CRC64;

Query Match 90.2%; Score 55; DB 1; Length 111;  
 Best Local Similarity 90.0%; Pred. No. 0.011;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PWTGCGTKL 10  
 DB 99 PWTGCGTKL 108

RESULT 7  
 ID KV3O\_MOUSE STANDARD; PRT; 111 AA.  
 AC P01667;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-III region PC 6308.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=79073152; PubMed=103003;  
 RA Weigert M., Gattalitan L., Loh E., Schilling J., Hood L.E.;  
 RT "Rearrangement of genetic information may produce immunoglobulin  
 RT diversity.";  
 RL Nature 276:785-790(1978).  
 RL PIR: C01937; KWS08.  
 DR HSSP; P01789; IMCP.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; Igv; 1.  
 KW Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.

```
FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 39 53 FRAMEWORK-2.
FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 61 92 FRAMEWORK-3.
FT DOMAIN 93 101 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 102 111 FRAMEWORK-4.
FT DISULFID 23 92 BY SIMILARITY.
FT NON_TER 111 111
SQ SEQUENCE 111 AA: 12071 MW: 744ADE4D6C256D29 CRC64;

Query Match
Best Local Similarity 90.2%; Score 55; DB 1; Length 111;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTFGSGTKL 10
    111111111
DB 99 PWTFGSGTKL 108

RESULT 8
KV3Q_MOUSE STANDARD; PRT; 111 AA.
AC P01659;
DT 21-JUL-1986 (Rel. 01, Last Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region PC 7769.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=79073152; PubMed=103003;
RA Weigert M., Galtmaitan L., Loh E., Schilling J., Hood L.E.;
RT "Rearrangement of genetic information may produce immunoglobulin
RT diversity.";
RT Nature 276:785-790(1978).
RU PIR: E01937; KVM569.
DR HSSP: P01789; IMCP.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig: 1.
DR SMART: SM00406; IGV: 1.
KM Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 39 53 FRAMEWORK-2.
FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 61 92 FRAMEWORK-3.
FT DOMAIN 93 101 FRAMEWORK-4.
FT DOMAIN 102 111 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 92 BY SIMILARITY.
FT NON_TER 111 111
SQ SEQUENCE 111 AA: 12011 MW: 6FAA345279356829 CRC64;

Query Match
Best Local Similarity 90.2%; Score 55; DB 1; Length 111;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTFGSGTKL 10
    111111111
DB 99 PWTFGSGTKL 108

RESULT 9
KV3G_MOUSE STANDARD; PRT; 112 AA.
AC P01659;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
```

```
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region TEPC 124.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RX MEDLINE=73140225; PubMed=4691517;
RA McKean D.J., Potter M., Hood L.E.;
RT "Mouse immunoglobulin chains. Pattern of sequence variation among
RT kappa chains with limited sequence differences.";
RT Biochemistry 12:760-771(1973).
DR PIR: A01933; KVM532.
DR HSSP: P01607; IREI.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig: 1.
DR SMART: SM00406; IGV: 1.
KM Immunoglobulin V region.
FT DOMAIN 1 23 FRAMEWORK-1.
FT DOMAIN 24 38 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 39 53 FRAMEWORK-2.
FT DOMAIN 54 60 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 61 92 FRAMEWORK-3.
FT DOMAIN 93 101 FRAMEWORK-4.
FT DOMAIN 102 111 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 23 92 BY SIMILARITY.
FT NON_TER 112 112
SQ SEQUENCE 112 AA: 12339 MW: 7CFD38DBE8E9D71 CRC64;

Query Match
Best Local Similarity 90.2%; Score 55; DB 1; Length 112;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTFGSGTKL 10
    111111111
DB 99 PWTFGSGTKL 108

RESULT 10
KV3F_MOUSE STANDARD; PRT; 132 AA.
AC P01658;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-II region MOPC 321 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-37.
RX MEDLINE=78235887; PubMed=98179;
RA Burstein Y., Schechter I.;
RT "Primary structures of N-terminal extra peptide segments linked to
RT the variable and constant regions of immunoglobulin light chain
RT precursors: implications on the organization and controlled
RT expression of immunoglobulin genes.";
RT Biochemistry 17:2392-2400(1978).
RN [2]
RP SEQUENCE OF 21-132.
RX MEDLINE=73140224; PubMed=4120629;
RA McKean D.J., Potter M., Hood L.E.;
RT "Mouse immunoglobulin chains. Partial amino acid sequence of a kappa
RT chain.";
RL Biochemistry 12:749-759(1973).
CC -I- MISCELLANEOUS: THE PARTIAL SEQUENCE OF THE C REGION OF THIS
CC BENCE-JONES PROTEIN WAS ALSO DETERMINED. IT DIFFERS FROM THAT
CC REPORTED FOR MOUSE MOPC 21 ONLY IN THE TRANSPOSITION OF TWO NEARBY
CC RESIDUES.
```

DATE RECEIVED

CC -1- MISCELLANEOUS:  
J. Immunol. 122:19  
RL  
RI

CC -1- MISCELLANEOUS:

CC BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
 DR PIR; B92808; KWS09.  
 DR HSSP; P01607; 1REI.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 KM Immunoglobulin V region.  
 FT DOMAIN 1 23  
 AC P01653; FRAMEWORK-1.  
 DT 21-JUL-1986 (Rel. 01, Created) COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 35 49 FRAMEWORK-2.  
 DT 21-JUL-1986 (Rel. 01, Last sequence update) COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 50 56 FRAMEWORK-3.  
 DE 15-JUL-1999 (Rel. 38, Last annotation update) FRAMEWORK-3.  
 DE Ig kappa chain V-V region J606. COMPLEMENTARITY-DETERMINING-3.  
 OS Mus musculus (Mouse). FRAMEWORK-4.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 FT DOMAIN 89 97  
 FT DISULFID 23 88 BY SIMILARITY.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11876 MW; 35C116BD60F79310 CRC64;

Query Match 85.2%; Score 52; DB 1; Length 108;  
 Best Local Similarity 90.0%; Pred. No. 0.035;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFGGGTKL 10  
 :|:|||||  
 DB 95 PYTFGGGTKL 104

RESULT 14  
 KVS\_MOUSE STANDARD; PRT; 108 AA.  
 ID KVS\_MOUSE  
 AC P01652;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-V region J606.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=82099361; PubMed=6798111;  
 RA Johnson N., Stankard J., Paul L., Hood L.;  
 RT "The complete V domain amino acid sequences of two myeloma inulin-binding proteins."  
 RL J. Immunol. 128:302-307(1982).  
 CC -I- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).  
 CC PIR; A92811; KWS06.  
 DR HSSP; P01607; 1REI.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00406; IGV; 1.  
 KM Immunoglobulin V region.  
 FT DOMAIN 1 23 FRAMEWORK-1.  
 AC P01653; COMPLEMENTARITY-DETERMINING-1.  
 DT 21-JUL-1986 (Rel. 01, Created) COMPLEMENTARITY-DETERMINING-1.  
 FT DOMAIN 35 49 FRAMEWORK-2.  
 DT 21-JUL-1986 (Rel. 01, Last sequence update) COMPLEMENTARITY-DETERMINING-2.  
 FT DOMAIN 50 56 FRAMEWORK-3.  
 DE 15-JUL-1999 (Rel. 38, Last annotation update) FRAMEWORK-3.  
 DE Ig kappa chain V-V region J606. COMPLEMENTARITY-DETERMINING-3.  
 OS Mus musculus (Mouse). FRAMEWORK-4.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 FT DOMAIN 89 97  
 FT DISULFID 23 88 BY SIMILARITY.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11810 MW; 8DE4DD31076F2AFB CRC64;

Query Match 85.2%; Score 52; DB 1; Length 108;  
 Best Local Similarity 90.0%; Pred. No. 0.035;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFGGGTKL 10  
 :|:|||||  
 DB 95 PYTFGGGTKL 104

RESULT 15  
 KVS\_MOUSE STANDARD; PRT; 108 AA.  
 ID KVS\_MOUSE  
 AC P01653;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-V region W3082.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 FT DOMAIN 89 97  
 FT DISULFID 23 88 BY SIMILARITY.  
 FT NON\_TER 108 108  
 SQ SEQUENCE 108 AA; 11850 MW; C5C145DC376F30CD CRC64;

Query Match 85.2%; Score 52; DB 1; Length 108;  
 Best Local Similarity 90.0%; Pred. No. 0.035;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTFGGGTKL 10  
 :|:|||||  
 DB 95 PYTFGGGTKL 104

Search completed: August 20, 2002, 13:17:53  
 Job time: 262 sec





**THIS PAGE BLANK (USPTO)**



Db 91 PWTFGGGTRL 100

# RESULT 2

09JL84 PRELIMINARY; PRT: 107 AA.

AC 09JL84; PRELIMINARY; PRT: 107 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
 DE (FRAGMENT).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE=20448942; PubMed=10992488;  
 RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;  
 RT "T-Cell-dependent antibody response to the dominant epitope of  
 streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive  
 with cardiac myosin."  
 RL Infect.Immun. 68:5803-5808(2000).  
 EL EMBL: AF206022; AAF69320.1; -.  
 DR HSSP; P80362; IMTL.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IGV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 107  
 SQ SEQUENCE 107 AA; 11648 MW; ACF9B1253ACA1ESD CRC64;

## Query Match

Best Local Similarity 100.0%; Score 61; DB 11; Length 107;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PWTFGGGTRL 10  
 |||||  
 Db 95 PWTFGGGTRL 104

# RESULT 3

0920E6 PRELIMINARY; PRT: 109 AA.

AC 0920E6; PRELIMINARY; PRT: 109 AA.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PTERIN-MIMICKING ANTI-IDIOPE KAPPA CHAIN VARIABLE REGION  
 DE (FRAGMENT).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Atkin J.D., Iape A., Jennings I.G., Horaitis O., Cotton R.G.H.;  
 RT "Definition of the Idiotope of Pterin-Mimicking Antibodies Expressed  
 in Mammalian Cells."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF307938; AAI09422.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 109  
 SQ SEQUENCE 109 AA; 11943 MW; DAD3F98E05DD1501 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 61; DB 11; Length 109;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGGTRL 10  
 |||||  
 Db 95 PWTFGGGTRL 104

# RESULT 4

091WS9 PRELIMINARY; PRT: 233 AA.

AC 091WS9; PRELIMINARY; PRT: 233 AA.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL 25.8 KDA PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-COLON;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC013496; AAH13496.1; -.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT NON\_TER 233  
 SQ SEQUENCE 233 AA; 25781 MW; B1C184DA149A16EB CRC64;

## Query Match

Best Local Similarity 100.0%; Score 61; DB 11; Length 233;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PWTFGGGTRL 10  
 |||||  
 Db 114 PWTFGGGTRL 123

# RESULT 5

09JL82 PRELIMINARY; PRT: 104 AA.

AC 09JL82; PRELIMINARY; PRT: 104 AA.  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ANTI-MYOSIN IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
 DE (FRAGMENT).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE=20448942; PubMed=10992488;  
 RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;  
 RT "T-Cell-dependent antibody response to the dominant epitope of  
 streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive  
 with cardiac myosin."  
 RL Infect.Immun. 68:5803-5808(2000).  
 DR EMBL: AF206024; AAF69322.1; -.  
 DR HSSP; P01607; IREI.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IGV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 104  
 SQ SEQUENCE 104 AA; 11360 MW; 5DA8BBF05FOAIAE CRC64;

## Query Match

Best Local Similarity 85.2%; Score 52; DB 11; Length 104;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRL 10  
1:|||||  
DB 92 PYTGGGTRL 101

RESULT 6  
ID Q920E9 PRELIMINARY; PRT: 111 AA.  
AC Q920E9;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PTERIN-MIMICKING ANTI-IDIOPOPE KAPPA CHAIN VARIABLE REGION  
DE (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Atkin J.D., Tape A., Jennings I.G., Horaitis O., Cotton R.G.H.;  
RT "Definition of the Idiotope of Pterin-Mimicking Antibodies Expressed  
in Mammalian Cells";  
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF307935; AAL09419.1; -  
FT NON\_TER 1 111  
FT NON\_TER 1 111  
SQ SEQUENCE 111 AA; 12046 MW; 1E46988AA6858526 CRC64;

Query Match 85.2%; Score 52; DB 11; Length 111;  
Best Local Similarity 90.0%; Pred. No. 0.29;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRL 10  
1:|||||  
DB 99 PYTGGGTRL 108

RESULT 7  
ID Q925S9 PRELIMINARY; PRT: 127 AA.  
AC Q925S9;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE IMMUNOGLOBULIN LIGHT CHAIN (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RX MEDLINE=99306687; PubMed=10380019;  
RA Tripathi P.K., Qin H., Bhattacharya-Chatterjee M., Ceriani R.L.,  
Foon K.A., Chatterjee S.K.;  
RT "Construction and characterization of a chimeric fusion protein  
consisting of an anti-idiotypic antibody mimicking a breast cancer-  
associated antigen and the cytokine GM-CSF";  
RL Hydridoma 18:193-202(1999).  
DR EMBL: AF144721; AAK5120.1; -  
FT NON\_TER 127 127  
FT NON\_TER 127 127  
SQ SEQUENCE 127 AA; 13794 MW; 13F61BEBB981EAF5 CRC64;

Query Match 85.2%; Score 52; DB 11; Length 127;  
Best Local Similarity 90.0%; Pred. No. 0.33;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRL 10

DB 115 PYTGGGTRL 124  
1:|||||

RESULT 8  
ID Q90YF0 PRELIMINARY; PRT: 298 AA.  
AC Q90YF0;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CN 8 SCFV.  
GN CN 8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE=SPLEEN;  
RX MEDLINE=20183931; PubMed=10706631;  
RA Shinohara N., Demura T., Fukuda H.;  
RT "Isolation of a vascular cell wall-specific monoclonal antibody  
recognizing a cell polarity by using a phage display subtraction  
method";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:2585-2590(2000).  
DR EMBL: AB036341; BAA88633.1; -  
DR HSPF; P01607; 1REL.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; Ig; 2.  
DR SMART: SM00406; IgV; 2.  
SQ SEQUENCE 298 AA; 31867 MW; E0F96B8A17004317 CRC64;

Query Match 85.2%; Score 52; DB 11; Length 298;  
Best Local Similarity 90.0%; Pred. No. 0.79;  
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PWTGGGTRL 10  
1:|||||  
DB 267 PYTGGGTRL 276

RESULT 9  
ID Q90L80 PRELIMINARY; PRT: 114 AA.  
AC Q90L80;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MYOSIN-REACTIVE IMMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION  
DE (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=98277139; PubMed=9614934;  
RX Wu X., Liu B., Van der Merwe P.L., Kalls N.N., Berney S.M.,  
Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
fetus";  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
DR EMBL: AF035034; AAD56270.1; -  
DR HSPF; P80362; 1WTL.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_V.  
DR Pfam: PF00047; Ig; 1.  
DR SMART: SM00406; IgV; 1.  
FT NON\_TER 1 114  
FT NON\_TER 114 114

```

SQ SEQUENCE 114 AA; 12775 MW; 070E31E210D1CB01 CRC64;

Query Match
Best Local Similarity 80.0%; Score 50; DB 4; Length 114;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PWTGGGCTKL 10
    1111111111
DB 101 PWTGGGCTKV 110

RESULT 10
C98T13 PRELIMINARY; PRT; 109 AA.
AC Q98T13;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE IMUNOGLOBULIN LIGHT CHAIN VARIABLE REGION (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RT "Nucleotide sequences encoding the variable regions of monoclonal AB,
RT A9-11-5, directed against antigen of Hepatitis B virus.";
RL Submitted (JUL-2000) to the EMBL/Genbank/DDJ databases.
DR EMBL; AF287275; MAG03053.1; -.
DR HSSP; P01703; 7FAB.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003600; Ig_1like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00409; Ig_1.
DR SMART; SM00406; IGV_1.
DR SMART; SM00410; IGV_1like; 1.
FT NON_TER 109 1
FT 109 1
FT SEQUENCE 109 AA; 11554 MW; 4F91ED9D351B1E158 CRC64;

Query Match
Best Local Similarity 80.3%; Score 49; DB 11; Length 109;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTFGGCTKL 10
    1111111111
DB 97 WTFGGCTKL 105

RESULT 11
C98D84 PRELIMINARY; PRT; 130 AA.
AC Q98D84;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE 1810027001RIK PROTEIN.
GN 1810027001RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kociba H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007622; BAB25142.1; -.
DR HSSP; P01842; 7FAB.
DR MGD; MGI:1916426; 1810027001RIK.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGV_1.
DR PROSITE; PS00290; IGV_MHC; UNKNOWN_1.
DR SEQUENCE 130 AA; 14253 MW; 438197975E76E54 CRC64;

Query Match
Best Local Similarity 80.3%; Score 49; DB 11; Length 130;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTFGGCTKL 10
    1111111111
DB 13 WTFGGCTKL 21

RESULT 12
C99IV32 PRELIMINARY; PRT; 233 AA.
AC C99IV32;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE ADULT MALE SMALL INTESTINE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:2010203H01, FULL INSERT SEQUENCE (ADULT MALE SMALL INTESTINE
DE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY, CLONE:2010005I06, FULL
DE INSERT SEQUENCE) (ADULT MALE SMALL INTESTINE CDNA, RIKEN FULL-LENGTH
DE ENRICHED LIBRARY, CLONE:2010013G21, FULL INSERT SEQUENCE).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RA Adachi J., Aizawa K., Akehira S., Akiyama T., Aono H., Arai A.,
RA Arakawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,
RA Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T.,
RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,
RA Shibata K., Shibata Y., Shitagawa A., Shiraki T., Sogabe Y.,
RA Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,
RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/Genbank/DDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
```

RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nucleotide 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitanai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsumoto M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuda S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 DR EMBL: AK008421; BAB25659.1; -;  
 DR EMBL: AK008129; BAB25479.1; -;  
 DR EMBL: AK008232; BAB25546.1; -;  
 SQ SEQUENCE 233 AA; 24874 MW; C4E896D46AABD2C1 CRC64;  
 QY 2 WTEGCGTKL 10  
 Db 116 WTEGCGTKL 124  
 Query Match 80.3%; Score 49; DB 11; Length 233;  
 Best Local Similarity 88.9%; Pred. No. 1.9;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 RESULT 13  
 ID 09JUL74 PRELIMINARY; PRT; 99 AA.  
 AC 09JUL74;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ANTI-MYOSIN IMMUNOGLUBULIN LIGHT CHAIN VARIABLE REGION  
 (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RX MEDLINE=20448942; PubMed=10992488;  
 RA Malkiel S., Liao L., Cunningham M.W., Diamond B.;  
 RT "T-Cell-dependent antibody response to the dominant epitope of  
 streptococcal polysaccharide, N-acetyl-glucosamine, is cross-reactive  
 with cardiac myosin.";  
 RL Infect. Immun. 68:5803-5808(2000).  
 DR EMBL: AF206032; AAP69330.1; -;  
 DR HSP; P80362; IWL.

DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IGV; 1.  
 FT NON\_TER 1  
 FT NON\_TER 99  
 SQ SEQUENCE 99 AA; 10939 MW; 3B25D0E78453324 CRC64;  
 Query Match 77.0%; Score 47; DB 11; Length 99;  
 Best Local Similarity 90.0%; Pred. No. 1.7;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 PWTGCGTKL 10  
 Db 87 PWTGCGTKL 96  
 RESULT 14  
 ID 09JUL74 PRELIMINARY; PRT; 211 AA.  
 AC 09JUL74;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ADULT MALE KIDNEY CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,  
 CDONE:0610010P20, FULL INSERT SEQUENCE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,  
 RA Atakawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotoh K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,  
 RA Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,  
 RA Nishi K., Nomura K., Numazaki R., Ono M., Okazaki Y., Okido T.,  
 RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,  
 RA Shibata K., Shibata Y., Shinagawa A., Shiraki T., Sogabe Y.,  
 RA Suzuki H., Tagami T., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,  
 RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nucleotide 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=20530913; PubMed=11076861;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.,  
 RT "RIKEN Integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 DR EMBL: AK002514; BAB22154.1; -  
 SQ SEQUENCE 211 AA; 23182 MW; 1A5FFA0F8BA50163 CRC64;

Query Match 77.0%; Score 47; DB 11; Length 211;  
 Best Local Similarity 90.0%; Pred. No. 3.6;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 PWTFGGTRL 10  
 1 | | | | | | | | | |  
 Db 92 PWTFGGTRL 101

RESULT 15  
 Q99M37 PRELIMINARY; PRT; 238 AA.  
 AC Q99M37;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOHETICAL 26.3 KDA PROTEIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS  
 RC TISSUE. ;  
 RA Strausberg R.;  
 RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: BC002035; AAH02035.1; -  
 DR HSSP; P01679; 2EPJ.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR003597; Ig\_c1.  
 DR InterPro: IPR003600; Ig\_like.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_2.  
 DR SMART: SM00409; Ig; 2.  
 DR SMART: SM00407; IGc1; 1.  
 DR SMART: SM00406; IGV; 1.  
 DR SMART: SM00410; IG\_like; 1.  
 DR PROSITE: PS00290; IG\_MHC; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 238 AA; 26344 MW; FB2B06A0B801330A CRC64;

Query Match 75.4%; Score 46; DB 11; Length 238;  
 Best Local Similarity 80.0%; Pred. No. 5.9;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 PWTFGGTRL 10  
 1 | | | | | | | | | |  
 Db 119 PWTFGGTRL 128

Search completed: August 20, 2002, 13:17:27  
 Job time: 241 sec





PT Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell anergy for  
 PT treatment of immunological diseases  
 XX  
 PS Claim 24; Page 84; 111pp; English.  
 CC  
 CC This peptide comprises an epitope of the human cytokine receptor  
 CC common gamma (gc) chain (see AAW31646) that is recognised by  
 CC gc blocking agents of the invention. 5 such epitopes (see  
 CC AAW31650-54) have been identified. The invention provides  
 CC compositions and methods for inhibiting cytokine signalling using  
 CC gc chain blocking agents for the treatment of immunological  
 CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,  
 CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel  
 CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,  
 CC parasitic infection, graft vs. host disease or psoriasis. A  
 CC preferred gc blocking agent is Mab CP-B8 or its Fab fragment (see  
 CC also AAW31647-48).  
 CC  
 SQ Sequence 5 AA:  
 QY 1 FNEY 5  
 Db 1 fney 5  
 Query Match 100.0%; Score 28; DB 19; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2  
 AAB48710  
 ID AAB48710 standard; peptide; 13 AA.  
 XX  
 AC AAB48710;  
 XX  
 DT 09-MAR-2001 (first entry)  
 DE Human cytokine receptor subunit h-gamma A-B loop.  
 XX  
 KW Haematopoietic growth factor receptor; cytokine receptor; subunit;  
 KW CR domain; CRD; h-beta-c subunit; GM-CSF receptor; IL-3 receptor;  
 KW IL-5 receptor; agonist; antagonist; myeloid leukemia; cancer;  
 KW lymphocyte leukaemia; non-haematopoietic tumour; inflammatory disease;  
 KW asthma; rheumatoid arthritis; atherosclerosis; CR module; CRM; A-B loop.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200066632-A1.  
 PD  
 XX 09-NOV-2000.  
 XX  
 PF 01-MAY-2000; 2000WO-AU00394.  
 XX  
 PR 29-APR-1999; 99AU-0000053.  
 XX  
 PA (MEDV-) MEDVET SCI PTY LTD.  
 XX  
 PI D'Andrea R, Bagley C, Vadas MA;  
 XX  
 DR WPI; 2001-049688/06.  
 XX  
 PT New agonists or antagonists of haematopoietic growth factors for treating  
 PT myeloid and lymphocyte leukemias, tumors and acute and chronic  
 PT inflammation such as asthma, rheumatoid arthritis and atherosclerosis -  
 XX  
 PS Disclosure; Fig 2; 35pp; English.  
 CC  
 CC The invention relates to a novel agonist or antagonist of a haematopoietic  
 CC growth factor which is capable of binding a region of the CRD3 (cytokine  
 CC receptor domain 3) of the human GM-CSF receptor/IL-3R/IL-5R shared  
 CC subunit h-beta-c, or an equivalent CRD in a different haematopoietic

CC growth factor receptor or other cytokine receptor. Binding of the  
 CC agonist or antagonist has an effect on the interaction between CRD3 and  
 CC CR4 (or analogous domain) which mediates the agonist or antagonist  
 CC property. In particular, the agonist or antagonist binds to the A-B loop  
 CC or the E-F loop of CRD3 or its equivalent. The invention also relates to  
 CC a method for screening and isolating a haematopoietic growth factor agonist  
 CC or antagonist, and pharmaceutical compositions containing an agonist or  
 CC antagonist. The agonists are used for treating conditions currently  
 CC treated by GM-CSF (granulocyte macrophage-colony stimulating factor),  
 CC IL-3 (interleukin-3), IL-5, and other members of the family of  
 CC haematopoietic growth factors. Haematopoietic growth factor antagonists are  
 CC useful for treating diseases such as myeloid and lymphocyte leukemias,  
 CC tumours of non-haematopoietic origin and acute and chronic inflammatory  
 CC conditions (e.g., asthma, rheumatoid arthritis and atherosclerosis).  
 CC Sequences AAB48693-B48716 represent the A-B loops of a variety of  
 CC cytokine receptor subunits from human and mouse.  
 CC  
 SQ Sequence 13 AA:  
 QY 1 FNEY 5  
 Db 4 fney 8  
 Query Match 100.0%; Score 28; DB 22; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3  
 AAB23875  
 ID AAB23875 standard; peptide; 21 AA.  
 XX  
 AC AAB23875;  
 XX  
 DT 17-JAN-2001 (first entry)  
 DE Artificial sequence designed peptide #1.  
 XX  
 KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200055182-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 15-MAR-2000; 2000WO-JP01584.  
 XX  
 PR 15-MAR-1999; 99JP-0067917.  
 XX  
 PA (SANY) SANKYO CO LTD.  
 XX  
 PI Kawaguchi J, Serizawa N;  
 XX  
 DR WPI; 2000-602106/57.  
 XX  
 PT Chemically synthesizing a peptide by solid phase synthesis by adding  
 PT protective group to inactivate part of alpha amino group of an amino  
 PT acid group -  
 XX  
 PS Example 1; Page 8; 38pp; Japanese.  
 CC  
 CC The present invention describes a method for chemically synthesizing a  
 CC peptide by solid phase synthesis. The method comprises processing a  
 CC carrier resin to which the C-terminal amino acid derivative of the  
 CC target peptide has been preliminarily bonded is characterised by adding  
 CC a protective group to inactivate a part of alpha-amino groups of an  
 CC amino acid capable of undergoing a peptide extension reaction on the  
 CC carrier resin. The method is useful for synthesizing a peptide by solid  
 CC phase synthesis. Long chain peptides can be readily synthesised in with  
 CC a high success ratio without using the segment condensation method or  
 CC gene manipulations. The present sequence represents an amino acid

CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 21 AA;

Query Match 100.0%; Score 28; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FNVEX 5  
|||||  
DB 5 fnvey 9

RESULT 4  
AAB23886  
ID AAB23886 standard; peptide; 21 AA.  
XX  
AC AAB23886;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Artificial sequence designed peptide #10.  
XX  
KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR WPI; 2000-602106/57.

XX Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group  
XX  
PS Example 3; Page 15; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 21 AA;

Query Match 100.0%; Score 28; DB 21; Length 21;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5  
|||||  
DB 5 fnvey 9

RESULT 5

AAB23876  
ID AAB23876 standard; peptide; 31 AA.  
XX  
AC AAB23876;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Artificial sequence designed peptide #2.  
XX  
KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR WPI; 2000-602106/57.

XX Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group  
XX  
PS Example 1; Page 9; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 31 AA;

Query Match 100.0%; Score 28; DB 21; Length 31;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5  
|||||  
DB 15 fnvey 19

RESULT 6  
AAB23887  
ID AAB23887 standard; peptide; 41 AA.  
XX  
AC AAB23887;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Artificial sequence designed peptide #11.  
XX  
KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.

RESULT 5

PD 21-SEP-2000.  
XX  
PS 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR WPI; 2000-602106/57.  
XX  
PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group  
XX  
PS Example 3; Page 15-16; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 41 AA;

Query Match 100.0%; Score 28; DB 21; Length 41;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
| | | | |  
Db 25 fnvey 29

RESULT 7  
AAB23877  
ID AAB23877 standard; peptide: 51 AA.  
XX  
AC AAB23877;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Artificial sequence designed peptide #3.  
XX  
KM Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR WPI; 2000-602106/57.  
XX  
PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group

XX  
PS Example 1; Page 9; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 51 AA;

Query Match 100.0%; Score 28; DB 21; Length 51;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
| | | | |  
Db 35 fnvey 39

RESULT 8  
AAB23888  
ID AAB23888 standard; peptide: 51 AA.  
XX  
AC AAB23888;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Artificial sequence designed peptide #12.  
XX  
KM Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR WPI; 2000-602106/57.  
XX  
PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group  
XX  
PS Example 3; Page 16; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX

SQL Sequence 51 AA;

Query Match 100.0%; Score 28; DB 21; Length 51;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 35 fnvey 39

RESULT 9

AAB23889  
ID AAB23889 standard; peptide; 61 AA.

AC AAB23889;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide #13.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

OS Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PF 15-MAR-2000; 2000WO-JP01584.

PR 15-MAR-1999; 99JP-0067917.

PA (SANY ) SANKYO CO LTD.

PI Kawaguchi J, Serizawa N;

DR WPI; 2000-602106/57.

PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group -

PS Example 3; Page 16; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesizing a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.

XX  
SQ Sequence 61 AA;

Query Match 100.0%; Score 28; DB 21; Length 61;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 45 fnvey 49

RESULT 10

AAB23878  
ID AAB23878 standard; peptide; 71 AA.

XX  
AC AAB23878;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide #4.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

OS Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PF 15-MAR-2000; 2000WO-JP01584.

PR 15-MAR-1999; 99JP-0067917.

PA (SANY ) SANKYO CO LTD.

PI Kawaguchi J, Serizawa N;

DR WPI; 2000-602106/57.

PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group -

PS Example 1; Page 10; 38pp; Japanese.

CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesizing a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.

XX  
SQ Sequence 71 AA;

Query Match 100.0%; Score 28; DB 21; Length 71;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 55 fnvey 59

RESULT 11

AAB23885  
ID AAB23885 standard; peptide; 79 AA.

AC AAB23885;

DT 17-JAN-2001 (first entry)

DE Artificial sequence designed peptide SEQ ID NO:3.

KW Peptide synthesis; chemical synthesis; solid phase synthesis.

OS Synthetic.

PN WO200055182-A1.

PD 21-SEP-2000.

PF 15-MAR-2000; 2000MO-JP01584.  
 XX  
 PR 15-MAR-1999; 99JP-0067917.  
 XX  
 PA (SANY ) SANKYO CO LTD.  
 XX  
 PI Kawaguchi J, Serizawa N;  
 XX  
 DR WPI; 2000-602106/57.  
 XX  
 PT Chemically synthesizing a peptide by solid phase synthesis by adding  
 PT protective group to inactivate part of alpha amino group of an amino  
 PT acid group  
 PS  
 PS Example 3; Page 32-33; 38pp; Japanese.  
 XX  
 CC The present invention describes a method for chemically synthesizing a  
 CC peptide by solid phase synthesis. The method comprises processing a  
 CC carrier resin to which the C-terminal amino acid derivative of the  
 CC target peptide has been preliminarily bonded is characterised by adding  
 CC a protective group to inactivate a part of alpha-amino groups of an  
 CC amino acid capable of undergoing a peptide extension reaction on the  
 CC carrier resin. The method is useful for synthesizing a peptide by solid  
 CC phase synthesis. Long chain peptides can be readily synthesised in with  
 CC a high success ratio without using the segment condensation method or  
 CC gene manipulations. The present sequence represents an amino acid  
 CC peptide sequence given in an example from the present invention.  
 CC  
 SQ Sequence 79 AA;  
 XX  
 OY 1 FNVEY 5  
 |||||  
 Db 63 fnvey 67  
 XX  
 RESULT 12  
 AAB23890  
 ID AAB23890 standard; peptide; 79 AA.  
 XX  
 AC AAB23890;  
 XX  
 DT 17-JAN-2001 (first entry)  
 XX  
 DE Artificial sequence designed peptide #14.  
 XX  
 KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200055182-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 15-MAR-2000; 2000MO-JP01584.  
 XX  
 PR 15-MAR-1999; 99JP-0067917.  
 XX  
 PA (SANY ) SANKYO CO LTD.  
 XX  
 PI Kawaguchi J, Serizawa N;  
 XX  
 DR WPI; 2000-602106/57.  
 XX  
 PT Chemically synthesizing a peptide by solid phase synthesis by adding  
 PT protective group to inactivate part of alpha amino group of an amino  
 PT acid group  
 PS  
 PS Example 3; Page 17; 38pp; Japanese.

XX  
 CC The present invention describes a method for chemically synthesizing a  
 CC peptide by solid phase synthesis. The method comprises processing a  
 CC carrier resin to which the C-terminal amino acid derivative of the  
 CC target peptide has been preliminarily bonded is characterised by adding  
 CC a protective group to inactivate a part of alpha-amino groups of an  
 CC amino acid capable of undergoing a peptide extension reaction on the  
 CC carrier resin. The method is useful for synthesizing a peptide by solid  
 CC phase synthesis. Long chain peptides can be readily synthesised in with  
 CC a high success ratio without using the segment condensation method or  
 CC gene manipulations. The present sequence represents an amino acid  
 CC peptide sequence given in an example from the present invention.  
 CC  
 SQ Sequence 79 AA;  
 XX  
 OY 1 FNVEY 5  
 |||||  
 Db 63 fnvey 67  
 XX  
 RESULT 13  
 AAB69096  
 ID AAB69096 standard; Peptide; 79 AA.  
 XX  
 AC AAB69096;  
 XX  
 DT 23-APR-2001 (first entry)  
 XX  
 DE Cedar pollen allergen T cell epitope derived peptide SEQ ID NO:3.  
 XX  
 KW Japanese cedar; Cryptomeria japonica; cedar pollen allergen;  
 KW T cell epitope; antisuigipollinosis.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN JP2000327699-A.  
 XX  
 PD 28-NOV-2000.  
 XX  
 PF 15-MAR-2000; 2000JP-0071710.  
 XX  
 PR 15-MAR-1999; 99JP-0068316.  
 XX  
 PA (HAYB ) HAYASHIBARA SEIBUTSU KAGAKU.  
 XX  
 PI (SANY ) SANKYO CO LTD.  
 XX  
 DR WPI; 2001-185061/19.  
 XX  
 PT Novel peptide and its use -  
 XX  
 PS Claim 13; Page 32-33; 75pp; Japanese.  
 XX  
 CC The present invention describes a peptide, its complex, derivative or  
 CC its polymerizate, where the peptide (I) has a formula of:  
 CC alpha1-alpha2-alpha3-alpha4-alpha5-alpha6-alpha7  
 CC where alpha1 to alpha7 - amino acid sequence selected from amino acid  
 CC sequences ranging from 11-19 amino acids derived from T cell epitopes  
 CC derived from cedar (Japanese cedar - Cryptomeria japonica) pollen  
 CC allergens. The peptide can be used in an antisuigipollinosis agent.  
 CC AAF59011 to AAF59062 and AAB69094 to AAB69121 represent sequences used  
 CC in the exemplification of the present invention.  
 CC  
 SQ Sequence 79 AA;  
 XX  
 OY 1 FNVEY 5  
 |||||  
 Db 63 fnvey 67  
 XX  
 Query Match 100.0%; Score 28; DB 22; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5  
|||||

Db 63 fnvey 67

RESULT 14  
AAB23874  
ID AAB23874 standard; peptide; 81 AA.  
XX  
AC AAB23874;  
XX  
DT 17-JAN-2001 (first entry)  
XX  
DE Artificial sequence designed peptide SEQ ID NO:1.  
XX  
KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR MPI; 2000-602106/57.  
XX  
PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group  
XX  
XX  
PS Example 1; Page 31; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 81 AA;

Query Match 100.0%; Score 28; DB 21; Length 81;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5  
|||||

Db 65 fnvey 69

RESULT 15  
AAB23879  
ID AAB23879 standard; peptide; 81 AA.  
XX  
AC AAB23879;  
XX  
DT 17-JAN-2001 (first entry)  
XX

DE Artificial sequence designed peptide #5.  
XX  
KW Peptide synthesis; chemical synthesis; solid phase synthesis.  
XX  
OS Synthetic.  
XX  
PN WO200055182-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 15-MAR-2000; 2000WO-JP01584.  
XX  
PR 15-MAR-1999; 99JP-0067917.  
XX  
PA (SANY ) SANKYO CO LTD.  
XX  
PI Kawaguchi J, Serizawa N;  
XX  
DR MPI; 2000-602106/57.  
XX

PT Chemically synthesizing a peptide by solid phase synthesis by adding  
PT protective group to inactivate part of alpha amino group of an amino  
PT acid group  
XX  
XX  
PS Example 1; Page 10; 38pp; Japanese.  
XX  
CC The present invention describes a method for chemically synthesizing a  
CC peptide by solid phase synthesis. The method comprises processing a  
CC carrier resin to which the C-terminal amino acid derivative of the  
CC target peptide has been preliminarily bonded is characterised by adding  
CC a protective group to inactivate a part of alpha-amino groups of an  
CC amino acid capable of undergoing a peptide extension reaction on the  
CC carrier resin. The method is useful for synthesising a peptide by solid  
CC phase synthesis. Long chain peptides can be readily synthesised in with  
CC a high success ratio without using the segment condensation method or  
CC gene manipulations. The present sequence represents an amino acid  
CC peptide sequence given in an example from the present invention.  
XX  
SQ Sequence 81 AA;

Query Match 100.0%; Score 28; DB 21; Length 81;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5  
|||||

Db 65 fnvey 69

Search completed: August 20, 2002, 11:07:15  
Job time: 6100 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:33 ; Search time 100.84 Seconds  
(without alignments)  
4.764 Million cell updates/sec

Title: US-09-824-286-13

Perfect score: 28

Sequence: 1 FNVEY 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	149	2 D97218	probable membrane
2	28	100.0	176	2 S28944	ribosomal protein
3	28	100.0	176	2 S55970	ribosomal protein
4	28	100.0	186	2 A81375	GTP cyclohydrolase
5	28	100.0	259	1 CRB02	carbonate dehydrat
6	28	100.0	276	1 B64372	ferredoxin - Metha
7	28	100.0	299	2 T11070	NADH dehydrogenase
8	28	100.0	301	2 G96944	probable permease
9	28	100.0	308	2 S58995	NADH dehydrogenase
10	28	100.0	312	2 S01191	NADH dehydrogenase
11	28	100.0	313	2 B71390	NADH dehydrogenase
12	28	100.0	314	2 T12010	NADH dehydrogenase
13	28	100.0	314	2 T09813	NADH dehydrogenase
14	28	100.0	315	1 OXMSIM	NADH dehydrogenase
15	28	100.0	317	2 T11337	NADH dehydrogenase
16	28	100.0	318	1 OXB01M	NADH dehydrogenase
17	28	100.0	318	1 OQRTIM	NADH dehydrogenase
18	28	100.0	318	2 A58850	NADH dehydrogenase
19	28	100.0	318	2 A58850	NADH dehydrogenase
20	28	100.0	318	2 S41835	NADH dehydrogenase
21	28	100.0	318	2 S47870	NADH dehydrogenase
22	28	100.0	318	2 S25151	NADH dehydrogenase
23	28	100.0	318	2 S41820	NADH dehydrogenase
24	28	100.0	318	2 A58888	NADH dehydrogenase
25	28	100.0	318	2 T11441	NADH dehydrogenase
26	28	100.0	318	2 T11493	NADH dehydrogenase
27	28	100.0	318	2 T11389	NADH dehydrogenase
28	28	100.0	318	2 T11857	NADH dehydrogenase
29	28	100.0	318	2 T10972	NADH dehydrogenase
				2 T11050	NADH dehydrogenase

30	28	100.0	318	2 T11480	NADH dehydrogenase
31	28	100.0	318	2 T11140	NADH dehydrogenase
32	28	100.0	318	2 T11247	NADH dehydrogenase
33	28	100.0	318	2 T11363	NADH dehydrogenase
34	28	100.0	318	2 T11428	NADH dehydrogenase
35	28	100.0	318	2 T45550	NADH dehydrogenase
36	28	100.0	319	2 T11454	NADH dehydrogenase
37	28	100.0	321	2 S55004	NADH dehydrogenase
38	28	100.0	321	2 T11274	NADH dehydrogenase
39	28	100.0	321	2 T13811	NADH dehydrogenase
40	28	100.0	322	2 S68128	NADH dehydrogenase
41	28	100.0	322	2 A90620	NADH dehydrogenase
42	28	100.0	322	2 A99626	NADH dehydrogenase
43	28	100.0	322	2 T11101	NADH dehydrogenase
44	28	100.0	322	2 T11127	NADH dehydrogenase
45	28	100.0	323	1 OXXLM	NADH dehydrogenase

## ALIGNMENTS

RESULT 1  
D97218  
probable membrane protein CAC2586 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: D97218  
R.Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; I  
J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: D97218  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-149 <KUR>  
A:Cross-References: GB:AE001437; PIDN:AAK80535.1; PID:g15025610; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2586

Query Match 100.0%; Score 28; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
Db 67 FNVEY 71

RESULT 2  
S28944  
ribosomal protein L6.e.A, cytosolic - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YML073c; ribosomal protein YL16.A  
C:Species: Saccharomyces cerevisiae  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
C:Accession: S28944; S78086  
R.Hashimoto, T.; Suzuki, K.; Mizuta, K.; Otake, E.  
Biochim. Biophys. Acta 1132, 195-198, 1992  
A:Title: Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the two gene  
A:Reference number: S28944; MUID:93003323  
A:Accession: S28944  
A:Molecule type: DNA  
A:Residues: 1-176 <HAS>  
A:Cross-References: EMBL:D10225; NID:g218508; PIDN:BAA01077.1; PID:g218509  
R.Brown, D.; Bowman, S.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S48816  
A:Accession: S78086  
A:Molecule type: DNA  
A:Residues: 1-176 <BRO>  
A:Cross-References: EMBL:Z46373; NID:g587529; PID:g914877; MIPS:YML073c

C:Genetics:  
 A:Gene: SGD:Y1L6A  
 A:Cross-references: MIPS:YML073c; SGD:S0004538  
 A:Map position: 13L  
 A:Introns: 5/3  
 A:Note: YML073c  
 C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 100.0%; Score 28; DB 2; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
 |||||  
 Db 101 FNVEY 105

RESULT 3  
 S5370  
 ribosomal protein L6.e.B, cytosolic - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein L9324.4; protein YLR448w; ribosomal protein YL16.B  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 21-Jul-2000  
 C:Accession: S55370; S28945; S11257  
 R:Du, Z.  
 submitted to the EMBL Data Library, March 1995  
 A:Description: The sequence of S. cerevisiae cosmid 9324.  
 A:Reference number: S55966  
 A:Accession: S55370  
 A:Molecule type: DNA  
 A:Residues: 1-176 <D0Z>  
 A:Cross-references: GB:022382; NID:g717059; PID:g717063; MIPS:YLR448w  
 A:Experimental sources: strain S288C (AB972)  
 R:Hashimoto, T.; Suzuki, K.; Mizuta, K.; Otake, E.  
 Biochim. Biophys. Acta 1132, 195-198, 1992  
 A:Title: Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the two genes  
 A:Reference number: S28944; MUID:93003323  
 A:Accession: S28945  
 A:Molecule type: DNA  
 A:Residues: 1-67, 'L', '69-176 <HNS>  
 A:Cross-references: EMBL:D10226; NID:g218510; PIDN:BAA01078.1; PID:g218511  
 R:Otake, E.; Higo, K.I.; Itoh, T.  
 Mol. Gen. Genet. 195, 544-546, 1984  
 A:Title: Yeast ribosomal proteins. VIII. Isolation of two proteins and sequence character  
 A:Reference number: S11249  
 A:Accession: S11257  
 A:Molecule type: protein  
 A:Residues: 2-8, 'X', '10-12, '2B', '15-25 <OTR>  
 C:Genetics:  
 A:Gene: SGD:RPL16B; YL16B  
 A:Cross-references: SGD:S0004440; MIPS:YLR448w  
 A:Map position: 12R  
 A:Introns: 5/3  
 C:Keywords: cytosol; protein biosynthesis; ribosome  
 F:2-176/Product: ribosomal protein L6.e #status experimental <MAT>

Query Match 100.0%; Score 28; DB 2; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
 |||||  
 Db 101 FNVEY 105

RESULT 4  
 AB1375  
 GTP cyclohydrolase II (EC 3.5.4.25) Cj0996 [imported] - Campylobacter jejuni (strain NCT  
 C:Species: Campylobacter jejuni  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
 C:Accession: AB1375

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Ch11  
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
 A:Reference number: AB1250; MUID:20150912  
 A:Accession: AB1375  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-186 <PAR>  
 A:Cross-references: GB:AL139076; GB:AL11168; NID:96968128; PIDN:CAB73252.1; PID:9696  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: rba; Cj0996  
 C:Superfamily: Escherichia coli cyclohydrolase II; cyclohydrolase homology  
 C:Keywords: hydrolase

Query Match 100.0%; Score 28; DB 2; Length 186;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
 |||||  
 Db 170 FNVEY 174

RESULT 5  
 CR802  
 carbonate dehydratase (EC 4.2.1.1) II - bovine (tentative sequence)  
 N:Alternate names: carbonic anhydrase II  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 31-Mar-1981 #sequence\_revision 31-Mar-1981 #text\_change 31-Mar-2000  
 C:Accession: A01144  
 R:Sciaky, M.; Limozin, N.; Filippi-Foveau, D.; Guillan, J.M.; Laurent-Tabasse, G.  
 Biochimie 58, 1071-1082, 1976  
 A:Title: Structure primaire de l'anhydrase carbonique erythrocytaire bovine CI. II.  
 A:Reference number: A90669; MUID:77065798  
 A:Accession: A01144  
 A:Molecule type: protein  
 A:Residues: 1-259 <SCI>  
 R:Guilan, J.M.; Limozin, N.; Mallet, B.; Di Costanzo, J.; Charrel, M.  
 Biochimie 59, 293-302, 1977  
 A:Title: Independance genetique de deux formes de l'anhydrase carbonique erythrocytal  
 A:Reference number: A90672; MUID:77242599  
 A:Contents: annotation  
 A:Note: one minor and two major forms were isolated chromatographically. One of the m  
 C:Superfamily: carbonate dehydratase; carbonic anhydrase homology  
 C:Keywords: acetylated amino end; carbon-oxygen lyase; hydro-lyase; zinc  
 F:4-258/Domain: carbonic anhydrase homology <CAH>  
 F:1/Modified site: acetylated amino end (Ser) #status experimental  
 F:93,95,118/Binding site: zinc (His) #status predicted

Query Match 100.0%; Score 28; DB 1; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
 |||||  
 Db 65 FNVEY 69

RESULT 6  
 B64372  
 ferredoxin - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: B64372  
 R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak  
 R:Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodex,  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
A:Reference number: A64300; MUID:96337999  
A:Accession: B64372  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-276 <BU>  
A:Cross-References: GB:U67506; GB:L77117; NID:91591274; PIDN:AAB98569.1; PID:91591285; T  
C:Genetics:  
A:Map position: REV514290-513460  
A:Start codon: TTG  
C:Superfamily: cell division inhibitor related protein; ferredoxin 2[4Fe-4S] homology  
F:69-119/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 100.0%; Score 28; DB 1; Length 276;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5  
|||||  
Db 212 FNVEX 216

RESULT 7  
T11070  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - nematode (*Onchocerca volvulus*) m  
C:Species: mitochondrion *Onchocerca volvulus*  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 12-Nov-1999  
C:Accession: T11070  
R:Kiedle, E.M.; Higazi, T.; Unasch, T.R.  
Mol. Biochem. Parasitol. 95, 111-127, 1998  
A:Title: The mitochondrial genome of *Onchocerca volvulus*: Sequence, structure and phylog  
A:Reference number: Z17246; MUID:98434257  
A:Accession: T11070  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-299 <KED>  
A:Cross-References: EMBL:AF015193; NID:92735934; PID:92735942; PIDN:AAC61617.1  
A:Experimental source: strain forest  
C:Genetics:  
A:Genome: mitochondrion  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 299;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5  
|||||  
Db 206 FNVEX 210

RESULT 8  
G96944  
Probable permease [imported] - *Clostridium acetobutylicum*  
C:Species: *Clostridium acetobutylicum*  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: G96944  
R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clo*  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: G96944  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-301 <KUR>  
A:Cross-References: GB:AF001437; PIDN:AK78346.1; PID:915023214; GSPDB:GM00168  
A:Experimental source: *Clostridium acetobutylicum* ATCC824  
C:Genetics:  
A:Gene: CAC0366

Query Match 100.0%; Score 28; DB 2; Length 301;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5  
|||||  
Db 51 FNVEX 55

RESULT 9  
S58995  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - earthworm (*Lumbricus terrestr*  
C:Species: mitochondrion *Lumbricus terrestris* (common earthworm)  
C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 07-Dec-1999  
C:Accession: S58995  
R:Boore, J.L.; Brown, W.M.  
Genetics 141, 305-319, 1995  
A:Title: Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus ter*  
A:Reference number: S58985; MUID:96042914  
A:Accession: S58995  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-308 <BOO>  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC4  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 100.0%; Score 28; DB 2; Length 308;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5  
|||||  
Db 211 FNVEX 215

RESULT 10  
S01191  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - fruit fly (*Drosophila melanoga*  
C:Species: mitochondrion *Drosophila melanogaster*  
C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 02-Nov-2001  
C:Accession: S01191  
R:Carssse, R.  
Genetics 118, 649-663, 1988  
A:Title: *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionar  
A:Reference number: S01185; MUID:88212147  
A:Accession: S01191  
A:Molecule type: DNA  
A:Residues: 1-312 <GAR>  
A:Note: the author translated the codon GTA for residue 190 as Ser and TCT for residu  
C:Genetics:  
A:Gene: ND-1  
A:Cross-References: FlyBase:FBgn0013679  
A:Genome: mitochondrion  
A:Genetic code: SGC4  
A:Start codon: ATA  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 100.0%; Score 28; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEX 5  
|||||

Db 215 FNVEY 219

RESULT 11

B71390

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - common lancelet mitochondrion

C:Species: mitochondrion Branchiostoma lanceolatum (common lancelet)

C>Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 20-Jun-2000

C:Accession: B71390

R:Spruyt, N.; Delarbre, C.; Gachejin, G.; Laudet, V.

A:Title: Complete sequence of the amphioxus (Branchiostoma lanceolatum) mitochondrial gene for NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1

A:Reference number: A71390; MUID:98292550

A:Accession: B71390

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-313 <SPR>

A:Cross-references: GB:Y16474; NID:g3292989; PIDN:CAA76247.1; PID:g3292991

C:Genetics:

A:Gene: NADH1

A:Genome: mitochondrion

A:Genetic code: SGC4

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 313;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5

|||||

Db 213 FNVEY 217

RESULT 12

T12010

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Anopheles quadrimaculatus A mitochondrion

C:Species: mitochondrion Anopheles quadrimaculatus A

C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-Jan-2000

C:Accession: T12010

R:Cockburn, A.F.; Mitchell, S.E.; Seawright, J.A.

A:Title: Insect Blochem. Physiol. 14, 31-36, 1990

A:Title: Cloning of the mitochondrial genome of Anopheles quadrimaculatus.

A:Reference number: Z17375; MUID:92190510

A:Accession: T12010

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-314 <COC>

A:Cross-references: EMBL:U04272; NID:g342501; PID:g1217663; PIDN:AAA93552.1

A:Experimental source: strain Orlando

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 314;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5

|||||

Db 216 FNVEY 220

RESULT 13

T09813

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - African malaria mosquito mitochondrion

C:Species: mitochondrion Anopheles gambiae (African malaria mosquito)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Oct-2000

C:Accession: T09813

R:Beard, C.B.; Hamm, D.M.; Collins, F.H.

Insect Mol. Biol. 2, 103-124, 1993

A:Title: The mitochondrial genome of the mosquito Anopheles gambiae: DNA sequence, gene organization, and phylogenetic relationships

A:Reference number: Z16863; MUID:97242550

A:Accession: T09813

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-314 <BEA>

A:Cross-references: EMBL:L20934; NID:g309056; PIDN:AAI2202.1; PID:g309069; GSPDB:GNO

A:Experimental source: strain G13

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC4

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 314;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5

|||||

Db 216 FNVEY 220

RESULT 14

OXMSIM

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - mouse mitochondrion

N:Alternate names: NADH-ubiquinone oxidoreductase chain 1

C:Species: mitochondrion Mus musculus (house mouse)

C>Date: 02-Apr-1982 #sequence\_revision 17-Jul-1998 #text\_change 07-Dec-1999

C:Accession: A00409

R:Bibb, M.J.; Van Etten, R.A.; Wright, C.T.; Walberg, M.W.; Clayton, D.A.

A:Title: Sequence and gene organization of mouse mitochondrial DNA.

A:Reference number: A00153; MUID:82137051

A:Accession: A00409

A:Molecule type: DNA

A:Residues: 1-315 <BBB>

A:Cross-references: GB:J01420; NID:g342520; PIDN:AA84644.1; PID:g896295

A:Note: the authors translated the initiation codon ATT for residue 1 as Ile

C:Genetics:

A:Gene: ND1

A:Genome: mitochondrion

A:Genetic code: SGC1

A:Start codon: ATT

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 1; Length 315;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5

|||||

Db 208 FNVEY 212

RESULT 15

T11337

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - aardvark mitochondrion

C:Species: mitochondrion Oryzteropus afer (aardvark)

C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

C:Accession: T11337

R:Arnasson, U.; Gullberg, A.; Janke, A.

Proc. R. Soc. Lond. B Biol. Sci. 266, 339-345, 1999

A:Title: The mitochondrial DNA molecule of the aardvark, Oryzteropus afer, and the po

A:Reference number: Z17263; MUID:99197468

A:Accession: T11337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-317 <ARN>

A:Cross-references: EMBL:Y18475; NID:g4691353; PIDN:CAB41621.1; PID:g4691354  
C:Genetics:  
A:Genome: mitochondrion  
A:Genetic code: SGC1  
A:Note: NADH1  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 100.0%; Score 28; DB 2; Length 317;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNVEY 5  
11111  
Db 211 FNVEY 215

Search completed: August 20, 2002, 11:10:34  
Job time: 5664 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:32:59 ; Search time 55.29 Seconds

(without alignments)  
3.501 Million cell updates/sec

Title: US-09-824-286-13

Perfect score: 28

Sequence: 1 FNVEY 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	175	1	RL6A_YEAST
2	28	100.0	175	1	RL6B_YEAST
3	28	100.0	259	1	CAH2_BOVIN
4	28	100.0	276	1	Y578_METJA
5	28	100.0	301	1	NDM1_NYCNO
6	28	100.0	303	1	NDM1_CABIN
7	28	100.0	303	1	NDM1_ZAGBR
8	28	100.0	308	1	NDM1_LUMTE
9	28	100.0	312	1	NDM1_DROME
10	28	100.0	312	1	NDM1_DROSU
11	28	100.0	313	1	NDM1_BRALA
12	28	100.0	314	1	NDM1_ANOGA
13	28	100.0	314	1	NDM1_ANOCU
14	28	100.0	315	1	NDM1_MOUSE
15	28	100.0	318	1	NDM1_BALMU
16	28	100.0	318	1	NDM1_BALPH
17	28	100.0	318	1	NDM1_BOVIN
18	28	100.0	318	1	NDM1_BRAVA
19	28	100.0	318	1	NDM1_CANFA
20	28	100.0	318	1	NDM1_CERST
21	28	100.0	318	1	NDM1_DASNO
22	28	100.0	318	1	NDM1_DIDMA
23	28	100.0	318	1	NDM1_EQUAS
24	28	100.0	318	1	NDM1_HALGR
25	28	100.0	318	1	NDM1_HIPAM
26	28	100.0	318	1	NDM1_HORSE
27	28	100.0	318	1	NDM1_LEMCA
28	28	100.0	318	1	NDM1_MACED
29	28	100.0	318	1	NDM1_MACRO
30	28	100.0	318	1	NDM1_MACRU
31	28	100.0	318	1	NDM1_MANTE
32	28	100.0	318	1	NDM1_MONDO
33	28	100.0	318	1	NDM1_NOTTY

34	28	100.0	318	1	NDM1_NYCNO	078697 nycticebus
35	28	100.0	318	1	NDM1_ORNAN	037717 ornithorhyn
36	28	100.0	318	1	NDM1_PERCU	078710 peromyscus g
37	28	100.0	318	1	NDM1_PHACT	078706 phascogale
38	28	100.0	318	1	NDM1_PHOVI	000505 phoca vitul
39	28	100.0	318	1	NDM1_PIG	079874 sus scrofa
40	28	100.0	318	1	NDM1_RABIT	079427 oryctolagus
41	28	100.0	318	1	NDM1_RAF	063889 rattus norv
42	28	100.0	318	1	NDM1_RHUN	096189 rhinoceros
43	28	100.0	318	1	NDM1_SARHA	078711 sarcophilus
44	28	100.0	318	1	NDM1_SHEEP	078747 ovis aries
45	28	100.0	318	1	NDM1_SWIMA	078712 smilnopsids

## ALIGNMENTS

```

RESULT 1
ID      RL6A_YEAST      STANDARD:      PRT: 175 AA.
AC      002326;
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      60S ribosomal protein L6-A (L17) (YL16) (RP18).
GN      RL6A OR YL16A OR YML073C.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93003323; PubMed=1390890;
RA      Hashimoto T., Suzuki K., Mizuta K., Otake E.;
RT      "Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the
RL      two genes encoding Saccharomyces cerevisiae YL16.";
RN      Biochim. Biophys. Acta 1132:195-198(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / AB972;
RA      Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL      Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC      -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L6 IN YEAST.
CC      -1- SIMILARITY: BELONGS TO THE L6E FAMILY OF RIBOSOMAL PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: D10225; BAA01077.1; -
CC      EMBL: Z46373; CAAB6505.1; -
CC      PIR: S28944; S28944.
CC      SGD: S0004538; RPL6A.
CC      InterPro: IPR000915; Ribosomal_L6E.
CC      Pfam: PF01159; Ribosomal_L6E; 1.
CC      ProDom: PD009612; Ribosomal_L6E; 1.
CC      PROSITE: PS01170; RIBOSOMAL_L6E; 1.
CC      KMW: Ribosomal protein; Multigene family.
CC      INIT MET 0
CC      SQUENCE 175 AA; 19830 MW; 0388034AAAF73BCB4 CRC64;

```

Query Match 100.0%; Score 28; DB 1; Length 175;

Best Local Similarity 100.0%; Pred. No. 7.1;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
DB 100 FNVEY 104

```

RESULT 2
ID RLB6_YEAST STANDARD: PRT: 175 AA.
AC P05739:
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L6-B (L17) (YLI6) (RP18).
GN RPL6B OR YLI6B OR RPL16B OR YLR448W OR L9324.4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93003323; PubMed=1390890;
RA Hashimoto T., Suzuki K., Mizuta K., Otake E.;
RT "Yeast ribosomal proteins: XIV. Complete nucleotide sequences of the
RL two genes encoding Saccharomyces cerevisiae YLI6."
RL Biochim. Biophys. Acta 1132:195-198(1992).
FN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Merdis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Taich A., Trevasis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-24.
RA Otake E., Higo K.-I., Itoh T.;
RT "Yeast ribosomal proteins: VIII. Isolation of two proteins and
RL sequence characterization of twenty-four proteins from cytoplasmic
RT ribosomes."
RL Mol. Gen. Genet. 195:544-546(1984).
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L6 IN YEAST.
CC -1- SIMILARITY: BELONGS TO THE L6E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10226; BAA01078.1; -
DR EMBL: U23282; AAB67529.1; -
DR PIR: S11257; S11257.
DR PIR: S28945; S28945.
DR SGD: S0004440; RPL6B.
DR InterPro: IPR000915; Ribosomal_L6E.
DR Pfam: PF01159; Ribosomal_L6e; 1.
DR ProDom: PD009612; Ribosomal_L6E; 1.
DR PROSITE: PS01170; RIBOSOMAL_L6E; 1.
KM Ribosomal protein; Multigene family.
FT INIT_MET 0
FT CONFLICT 67 67 P -> L (IN REF. 1).
SQ SEQUENCE 175 AA; 19855 MW; 689AFA0813DC6651 CRC64;

```

Query Match 100.0%; Score 28; DB 1; Length 175;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 Db 100 FNVEY 104

```

RESULT 3
ID CAH2_BOVIN STANDARD: PRT: 259 AA.
AC P00921:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbonic anhydrase II (EC 4.2.1.1) (Carbonate dehydratase II) (CA-II).
GN CA2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Erythrocyte;
RX MEDLINE=77065798; PubMed=826282;
RA Sclaky M., Limozin N., Filippi-Foveau D., Guillan J.M.,
RA Laurent-Tabuse G.;
RT "Primary structure of bovine erythrocyte carbonic anhydrase
RT Cl. II. Complete sequence."
RL Biochimie 58:1071-1082(1976).
RN [2]
RP REVISIONS.
RX MEDLINE=77242599; PubMed=19093;
RA Guillan J.M., Limozin N., Mallet B., di Costanzo J., Charrel M.;
RT "Genetic independence of two forms of carbonic anhydrase from bovine
RT erythrocytes."
RL Biochimie 59:293-302(1977).
CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- MISCELLANEOUS: ONE MINOR AND TWO MAJOR FORMS WERE ISOLATED
CC CHROMATographically.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC FAMILY.
CC -1- DATABASE: NAME=worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/CA.html".
DR PIR: A01144; CR02.
DR HSP: P00918; IBV3.
DR InterPro: IPR001148; Carb_anhydrase.
DR Pfam: PF00194; Carb_anhydrase; 1.
DR ProDom: PD000865; Carb_anhydrase; 1.
DR PROSITE: PS00162; EUK_CO2_ANNHYDRASE; 1.
KM Lyase; Zinc; Acetylation.
FT MOD_RES 1 1 ACETYLATION.
FT METAL 93 93 ZINC (CATALYTIC).
FT METAL 95 95 ZINC (CATALYTIC).
FT METAL 118 118 ZINC (CATALYTIC).
FT VARIANT 56 56 R -> Q (IN ONE OF THE MAJOR FORMS).
SQ SEQUENCE 259 AA; 28980 MW; 8488644617BF32D8 CRC64;

```

Query Match 100.0%; Score 28; DB 1; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 Db 65 FNVEY 69

```

RESULT 4
ID Y578_METJA STANDARD: PRT: 276 AA.
AC Q57998:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M0578.

```



GN MJ0578. Methanococcus jannaschii.  
OS Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.T.,  
RA Overleak R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
RT Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii. "  
RT Science 273:1058-1073(1996).  
RL -1- SIMILARITY: BELONGS TO THE BACTERIAL TYPE FERREDOXIN FAMILY.  
CC -1- SIMILARITY: TO M.JANNASCHII MJ0579.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U67506; AAB98569.1; -  
CC HSSP: P00198; IFCA.  
DR TIGR: MJ0578; -  
DR InterPro: IPR001450; 4Fe4S-ferredoxin.  
DR InterPro: IPR000707; Para.  
DR Pfam: PF00037; fer4; 2.  
DR PROSITE: PS00198; 4Fe4S-FERREDOXIN; 1.  
DR Hypothetical protein; Iron-sulfur; 4Fe-4S; ATP-binding;  
KM Complete proteome.  
FT NP\_BIND 15 22 ATP (POTENTIAL);  
FT METAL 76 76 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
FT METAL 79 79 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
FT METAL 82 82 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
FT METAL 86 86 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
FT METAL 101 101 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
FT METAL 104 104 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
FT METAL 107 107 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).  
FT METAL 111 111 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 276 AA; 30794 MW; 0F900CE7F2E88669 CRC64;  
  
Query Match 100.0%; Score 28; DB 1; Length 276;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FNVEY 5  
DB 212 FNVEY 216  
  
RESULT 5  
NUM1\_NYCNO STANDARD; PRT; 301 AA.  
AC 080003;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3) (Fragment).  
GN MTND1 OR NDI OR NADH1.  
OS Myctalus noctula (Noctule).  
OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Vespertilionidae;  
OC Myctalus.  
OX NCBI\_TaxID=51300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Petit E., Excoffier L., Mayer F.;  
RT "No evidence of bottlenecks in the post-glacial recolonization of  
RT Europe by the noctule bat (*Myctalus noctula*). "  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBD databases.  
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF065104; AAC28345.1; -  
CC EMBL: AF065108; AAC28349.1; -  
CC EMBL: AF065109; AAC28350.1; -  
DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
DR Pfam: PF00146; NADHdh; 1.  
DR PROSITE: PS00667; COMPLEXI\_ND1\_1; 1.  
DR PROSITE: PS00668; COMPLEXI\_ND1\_2; 1.  
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
FT NON\_TER 1 301  
SQ SEQUENCE 301 AA; 33957 MW; B294975131BB0D09 CRC64;  
  
Query Match 100.0%; Score 28; DB 1; Length 301;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FNVEY 5  
DB 205 FNVEY 209  
  
RESULT 6  
NUM1\_CABUN STANDARD; PRT; 303 AA.  
AC 078700;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3) (Fragment).  
GN MTND1 OR NDI.  
OS Cabassous unicinctus (Southern naked-tailed armadillo).  
OS Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Edentata; Dasypodidae; Cabassous.  
OX NCBI\_TaxID=48852;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=98404151; PubMed=9732458;  
RA Cao Y., Janke A., Waddell P.J., Westernman M., Takenaka O., Murata S.,  
RA Okada N., Paabo S., Hasegawa M.;  
RT phylogeny of eutherian orders. "  
RL J. Mol. Evol. 47:307-322(1998).  
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB011217; BAA32109.1; -  
 DR InterPro: IPR001694; Resp-chain\_NADH\_DHL.  
 DR Pfam: PF001146; NADHdh. 1.  
 DR PROSITE: PS00667; COMPLEX1\_NDL\_1; 1.  
 DR PROSITE: PS00668; COMPLEX1\_NDL\_2; 1.  
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 FT NON\_TER 1  
 FT NON\_TER 303  
 SQ SEQUENCE 303 AA; 33839 MW; 692D072836585532 CRC64;

Query Match 100.0%; Score 28; DB 1; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 Db 199 FNVEY 203

RESULT 7  
 NUIW\_ZAGBR STANDARD; PRT; 303 AA.  
 AC 078713;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3) (Fragment).  
 GN MNND1 OR NDI.  
 OS Zaglossus bruijnii (Long-beaked echidna).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Monotremata; Techylosidae; Zaglossus.  
 OX NCBI\_TaxID=33543;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98404151; PubMed=9732458;  
 RA Cao Y., Janke A., Muddell P.J., Westernman M., Takenaka O., Murata S.,  
 RA Orada N., Paabo S., Hasegawa M.;  
 RT "Conflict among individual mitochondrial proteins in resolving the  
 RT phylogeny of eutherian orders";  
 RL J. Mol. Evol. 47:307-322(1998).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB011230; BAA32122.1; -  
 DR InterPro: IPR001694; Resp-chain\_NADH\_DHL.  
 DR Pfam: PF001146; NADHdh. 1.  
 DR PROSITE: PS00667; COMPLEX1\_NDL\_1; 1.  
 DR PROSITE: PS00668; COMPLEX1\_NDL\_2; 1.  
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 FT NON\_TER 1  
 FT NON\_TER 303  
 SQ SEQUENCE 303 AA; 34223 MW; 64E8AF098E26FC9F CRC64;

Query Match 100.0%; Score 28; DB 1; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 Db 199 FNVEY 203

RESULT 8  
 NUIW\_LUMTE STANDARD; PRT; 308 AA.  
 AC Q37546;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 38, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
 GN NDI.  
 OS Lumbricus terrestris (Common earthworm).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 OC Lumbricina; Lumbricidae; Lumbricus.  
 OX NCBI\_TaxID=6398;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96042914; PubMed=8536978;  
 RA Boore J.L., Brown W.M.;  
 RT "Complete sequence of the mitochondrial DNA of the annelid worm  
 RT Lumbricus terrestris";  
 RL Genetics 141:305-319(1995).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U24570; AAC46874.1; -  
 DR InterPro: IPR001694; Resp-chain\_NADH\_DHL.  
 DR Pfam: PF001146; NADHdh. 1.  
 DR PROSITE: PS00667; COMPLEX1\_NDL\_1; 1.  
 DR PROSITE: PS00668; COMPLEX1\_NDL\_2; 1.  
 KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 SQ SEQUENCE 308 AA; 34841 MW; 6FAE24B35D566DBC CRC64;

Query Match 100.0%; Score 28; DB 1; Length 308;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 Db 211 FNVEY 215

RESULT 9  
 NUIW\_DROME STANDARD; PRT; 312 AA.  
 AC P18929;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
 GN MT:ND1 OR NDI.  
 OS Drosophila melanogaster (Fruit fly).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRETAGNE;

```

RX MEDLINE=88212147; PubMed=3130291;
RA Garesse R.;
RT "Drosophila melanogaster mitochondrial DNA: gene organization and
  evolutionary considerations.";
RL Genetics 118:649-663(1988).
CC
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: M37275; AAA69715.1; -
DR EMBL: U37341; AAC47823.1; -
DR PIR: S01191; S01191.
DR FlyBase: FBgn0013679; mt:ND1.
DR InterPro: IPR001694; Resp_chain_NADH_DH1.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00668; COMPLEX1_ND1_2; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
KW
SQ SEQUENCE 312 AA; 35910 MW; B2B01BE6213CE722 CRC64;

Query Match          100.0%; Score 28; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FENVY 5
Db 215 FENVY 219

RESULT 10
ID NUIM_DROSU STANDARD; PRT; 312 AA.
AC P51937;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MT:ND1 OR ND1.
OS Drosophila subobscura (Fruit fly).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN 11
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=95106285; PubMed=7528808;
RA Barrio E., Latorre A., Moya A.;
RT "Phylogeny of the Drosophila obscura species group deduced from
  mitochondrial DNA sequences.";
RL J. Mol. Evol. 39:478-488(1994).
RN 12
RP SEQUENCE OF 50-312 FROM N.A.
RA STRAIN-SSP. TUE 3;
RA Volz-Lingenhl A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U07288; AAA76626.1; -
DR EMBL: X65130; CAA46260.1; -
DR FlyBase: FBgn0012956; Dsuh\mt:ND1.
DR InterPro: IPR001694; Resp_chain_NADH_DH1.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE: PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE: PS00668; COMPLEX1_ND1_2; 1.
DR Mitochondrion; Oxidoreductase; NAD; Transmembrane.
KW
SQ SEQUENCE 312 AA; 35877 MW; AFD1B3BC2BCD868 CRC64;

Query Match          100.0%; Score 28; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FENVY 5
Db 215 FENVY 219

RESULT 11
ID NUIM_BRALA STANDARD; PRT; 313 AA.
AC O21000; O47433;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN ND1 OR NAD1 OR NADH1.
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus), and
  Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7740; 7739;
RN 11
RP SEQUENCE FROM N.A.
RC SPECIES=B.lanceolatum;
RA MEDLINE=97398704; PubMed=9254918;
RA Delarbre C., Barriel V., Tillier S., Janvier P., Gachelin G.;
RT "The main features of the cranial mitochondrial DNA between the ND1
  and the COI genes were established in the common ancestor with the
  lancelet.";
RL Mol. Biol. Evol. 14:807-813(1997).
RN 12
RP SEQUENCE FROM N.A.
RC SPECIES=B.lanceolatum;
RA MEDLINE=98292550; PubMed=9628930;
RA Spruyt N., Delarbre C., Gachelin G., Laudet V.;
RT "Complete sequence of the amphioxus (Branchiostoma lanceolatum)
  mitochondrial genome: relations to vertebrates.";
RL Nucleic Acids Res. 26:3279-3285(1998).
RN 13
RP SEQUENCE FROM N.A.
RC SPECIES=B.floridae;
RA MEDLINE=99261652; PubMed=10331267;
RA Boore J.L., Daehler L.L., Brown W.M.;
RT "Complete sequence, gene arrangement, and genetic code of
  mitochondrial DNA of the cephalochordate Branchiostoma floridae
  (Amphioxus).";
RL Mol. Biol. Evol. 16:410-418(1999).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).

```

```
CC -----
DR EMBL: Y09524; CAA70708.1; -
DR EMBL: Y16474; CAA76247.1; -
DR EMBL: AF098298; AAB88002.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_2; 1.
DR OXidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT CONFLICT 111 111 S -> P (IN REF. 1).
FT CONFLICT 274 274 R -> S (IN REF. 1).
SQ SEQUENCE 313 AA; 34429 MW; 1787EA3049F0B959 CRC64;

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 313;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5
Db 213 FNVEX 217

RESULT 12
NUIM_ANOGA STANDARD; PRT; 314 AA.
AC P34846;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN NDI.
OS Anopheles gambiae (African malaria mosquito).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G3;
RX MEDLINE=97242550; PubMed=9087549;
RA Beard C.B., Hamm D.M., Collins F.H.;
RT "The mitochondrial genome of the mosquito Anopheles gambiae: DNA
RT sequence, genome organization, and comparisons with mitochondrial
RT sequences of other insects."
RL Insect Mol. Biol. 2:103-104(1993).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L20934; AAD12202.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_2; 1.
DR OXidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 314 AA; 36100 MW; 512CABE0AA4E1DB7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 314;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5
Db 213 FNVEX 217
```

```
Db 216 FNVEX 220

RESULT 13
NUIM_ANOOU STANDARD; PRT; 314 AA.
AC P33502;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN NDI.
OS Anopheles quadrimaculatus (Mosquito).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7166;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORLANDO;
RX MEDLINE=92190510; PubMed=2134168;
RA Cockburn A.F., Mitchell S.E., Seawright J.A.;
RT "Cloning of the mitochondrial genome of Anopheles quadrimaculatus."
RL Arch. Insect Biochem. Physiol. 14:31-36(1990).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04272; AAA93552.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADHdh; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR PROSITE; PS00667; COMPLEX1_ND1_2; 1.
DR OXidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 314 AA; 36138 MW; 7574D2C4C78E4F35 CRC64;

Query Match
Best Local Similarity 100.0%; Score 28; DB 1; Length 314;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEX 5
Db 216 FNVEX 220

RESULT 14
NUIM_MOUSE STANDARD; PRT; 315 AA.
AC P03888;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).
GN MTND1 OR MT-ND1 OR NDI.
OS Mus musculus (Mouse).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82137051; PubMed=7332926;
RA Bibb M.J., van Etten R.A., Wright C.T., Walberg M.W., Clayton D.A.;
RT "Sequence and gene organization of mouse mitochondrial DNA.";
```

RL Cell 26:167-180(1981).  
 CC -I- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: V00711; CAA24080.1; -;  
 DR EMBL: J01420; AAB48644.1; -;  
 DR PIR: A00409; OXMSIM.  
 DR MGD: MGI:101787; mt-ND1.  
 DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
 DR Pfam: PF00146; NADHdh; 1.  
 DR PROSITE: PS00667; COMPLEX1\_ND1\_1; 1.  
 DR PROSITE: PS00668; COMPLEX1\_ND1\_2; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.  
 SQ SEQUENCE 315 AA; 35651 MW; B5018F305AEC3B3A CRC64;

Query Match 100.0%; Score 28; DB 1; Length 315;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNVFY 5  
 |||||  
 DB 208 FNVFY 212

RESULT 15  
 NTJM\_BALMU  
 ID NTJM\_BALMU STANDARD; PRT; 318 AA.  
 AC P41296;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
 GN MTND1 OR ND1 OR NADH1.  
 OS Balanoptera musculus (blue whale).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Balanopteridae; Balanoptera.  
 OX NCBI\_TaxID=9771;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94141932; PubMed=8308901;  
 RA Arnason U., Gullberg A.;  
 RT "Comparison between the complete mtDNA sequences of the blue and the  
 RT fin whale, two species that can hybridize in nature."  
 RL J. Mol. Evol. 37:312-322(1993).  
 CC -I- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X72204; CAA50995.1; -;  
 DR PIR: S41820; S41820.  
 DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
 DR Pfam: PF00146; NADHdh; 1.  
 DR PROSITE: PS00667; COMPLEX1\_ND1\_1; 1.  
 DR PROSITE: PS00668; COMPLEX1\_ND1\_2; 1.  
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.

SQ SEQUENCE 318 AA; 35765 MW; 8BFB423C219BA57E CRC64;  
 Query Match 100.0%; Score 28; DB 1; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FNVFY 5  
 |||||  
 DB 211 FNVFY 215

Search completed: August 20, 2002, 11:33:00  
 Job time: 1449 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:52 ; Search time 191.14 Seconds  
(without alignments)  
4.525 Million cell updates/sec

Title: US-09-824-286-13

Perfect score: 28  
Sequence: 1 FNVEY 5

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rv1rus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	32	8 Q94VL5	Q94VL5 salmo trutt
2	28	100.0	32	8 Q94NY1	Q94NY1 salmo salar
3	28	100.0	80	10 Q9AW31	Q9AW31 guillardi
4	28	100.0	114	8 Q63938	Q63938 bufo gargar
5	28	100.0	114	8 Q63937	Q63937 bufo gargar
6	28	100.0	114	8 Q64007	Q64007 bufo andrew
7	28	100.0	114	8 Q63322	Q63322 bufo exsul
8	28	100.0	114	8 Q63324	Q63324 bufo viridi
9	28	100.0	114	8 Q63326	Q63326 bufo verric
10	28	100.0	114	8 Q63328	Q63328 bufo andrew
11	28	100.0	114	8 Q63330	Q63330 bufo gargar
12	28	100.0	119	8 Q9MDE1	Q9MDE1 coregonus a
13	28	100.0	119	8 Q9MDE1	Q9MDE1 oncorhynch
14	28	100.0	119	8 Q9MDE1	Q9MDE1 oncorhynch
15	28	100.0	121	8 Q958M0	Q958M0 rana catesb
16	28	100.0	121	8 Q958L7	Q958L7 rana sylvat

17	28	100.0	121	8 Q958L4	Q958L4 rana tempor
18	28	100.0	121	8 Q958L1	Q958L1 rana boylia
19	28	100.0	121	8 Q958K8	Q958K8 rana pretio
20	28	100.0	121	8 Q958K5	Q958K5 rana aurora
21	28	100.0	121	8 Q958K2	Q958K2 rana cascad
22	28	100.0	121	8 Q958J9	Q958J9 rana muscos
23	28	100.0	121	8 Q94PN0	Q94PN0 rana muscos
24	28	100.0	124	1 Q9P9G3	Q9P9G3 methanosarc
25	28	100.0	126	8 Q9BAX2	Q9BAX2 notophthalm
26	28	100.0	126	8 Q9B4W9	Q9B4W9 lytiocrito
27	28	100.0	126	8 Q9B4W6	Q9B4W6 pachytirito
28	28	100.0	126	8 Q9B4W3	Q9B4W3 triturus vu
29	28	100.0	126	8 Q9B4W0	Q9B4W0 chioglossa
30	28	100.0	126	8 Q9B4V7	Q9B4V7 merensieill
31	28	100.0	126	8 Q9B4V4	Q9B4V4 merensieill
32	28	100.0	126	8 Q9B4V2	Q9B4V2 salamandra
33	28	100.0	126	8 Q9B4V0	Q9B4V0 salamandra
34	28	100.0	126	8 Q9B4U7	Q9B4U7 merensieill
35	28	100.0	126	8 Q9B4U4	Q9B4U4 merensieill
36	28	100.0	126	8 Q9B4U1	Q9B4U1 merensieill
37	28	100.0	126	8 Q9B4T9	Q9B4T9 merensieill
38	28	100.0	126	8 Q9B4T7	Q9B4T7 merensieill
39	28	100.0	126	8 Q9B4T5	Q9B4T5 merensieill
40	28	100.0	126	8 Q9B4T3	Q9B4T3 merensieill
41	28	100.0	126	8 Q9B4T0	Q9B4T0 merensieill
42	28	100.0	126	8 Q9B4S7	Q9B4S7 merensieill
43	28	100.0	126	8 Q9B4S4	Q9B4S4 merensieill
44	28	100.0	126	8 Q9B4S2	Q9B4S2 merensieill
45	28	100.0	127	8 Q9TGA0	Q9TGA0 shinisaurus

#### ALIGNMENTS

RESULT 1  
Q94VL5 PRELIMINARY; PRT; 32 AA.  
AC Q94VL5;  
DT 01-DEC-2001 (TREMREL.19, Created)  
DT 01-DEC-2001 (TREMREL.19, Last sequence update)  
DT 01-DEC-2001 (TREMREL.19, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).  
GN ND1.  
OS Salmo trutta (Brown trout).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
OX NCBI\_TaxID=8032;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Arnold A.L., Knox D., Verspoor E.;  
RT "mtDNA sequence from brown trout from the River Dee, Scotland;  
RT Position 4447 to 4546.";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF405331; AAK97505.1; -;  
KW Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 32  
SQ SEQUENCE 32 AA; 3425 MW; 90180CC34D5F08AB CRC64;

Query Match 100.0%; Score 28; DB 8; Length 32;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
|||||  
Db 16 FNVEY 20

RESULT 2  
Q94NY1

```

ID 094NY1 PRELIMINARY; PRT; 32 AA.
AC 094NY1:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).
GN NDI.
OS Salmo salar (Atlantic salmon).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V59, AND V62;
RA Gonzalez-Morales M.R., Verspoor E.;
RT "Mitochondrial DNA variation in Paleolithic and modern Atlantic salmon
from the Iberian Glacial Refugium."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arnold A.L., Knox D., Verspoor E.;
RT "mtDNA sequence from Atlantic salmon from the Saint John River,
Canada: Position 4447 to 4546."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Arnold A.L., Knox D., Verspoor E.;
RT "mtDNA sequence from Atlantic salmon from the River Torne, Sweden:
Position 4447 to 4546."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF385752; AAK67347.1; -
DR EMBL; AF385753; AAK67348.1; -
DR EMBL; AF405329; AAK97503.1; -
DR EMBL; AF405330; AAK97504.1; -
KM Mitochondrion.
FT NON_TER 1 1
FT NON_TER 32 32
SQ SEQUENCE 32 AA; 3425 MW; 90180CC34D5F08AB CRC64;

Query Match 100.0%; Score 28; DB 8; Length 32;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 16 FNVEY 20

RESULT 3
O9AW31 PRELIMINARY; PRT; 80 AA.
ID O9AW31:
AC O9AW31:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 9.6 KDA PROTEIN.
OS Guillardia theta (Cryptomonas phi).
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20087226; PubMed=10618395;
RA Zanner S., Fraunholz M., Wastl J., Penny S.L., Beaton M.,
CA Cavalier-Smith T., Maier U., Douglas S.;
RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd
telomeres in an unusually compact eukaryotic genome, the cryptomonad
nucleomorph."
RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zanner S., Fraunholz M., Beaton M., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.;
RT "The highly reduced genome of an enslaved algal nucleus."
RL Nature 410:1091-1096(2001).
DR EMBL; AJ010592; CAC27040.1; -
KM Hypothetical protein.
SQ SEQUENCE 80 AA; 9620 MW; 592C0AEB96A457DD CRC64;

Query Match 100.0%; Score 28; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 23 FNVEY 27

RESULT 4
O63938 PRELIMINARY; PRT; 114 AA.
ID O63938:
AC O63938:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo gargarizans.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tunliye B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
from the eastern escarpment of the Tibetan Plateau: a case of
vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
DR EMBL; AF004535; AAD05148.1; -
DR EMBL; AF004533; AAD05144.1; -
DR EMBL; AF004534; AAD05146.1; -
DR InterPro: IPR001694; Resp_chain_NADH_DHL.
DR Pfam: PF00146; NADhdh; 1.
KM Mitochondrion.
FT NON_TER 1 1
FT NON_TER 114 114
SQ SEQUENCE 114 AA; 12892 MW; FB68884EC1B90E97 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5
Db 8 FNVEY 12

RESULT 5
O63937 PRELIMINARY; PRT; 114 AA.
ID O63937:
AC O63937:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 18, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).
GN NDI.
OS Bufo gargarizans.

```



OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
OX NCBL\_TaxID=30331;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RT Tunliye B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal.";  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
DR EMBL; AF004532; AAD05142.1; -;  
DR EMBL; AF004531; AAD05140.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NMDH.DH1.  
DR Pfam; PF00146; NADHdh; 1.  
KW Mitochondrion.  
FT NON\_TER  
SQ SEQUENCE 114 AA; 12844 MW; EF3884EDAA20E8C CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNV5 5  
|  
|  
|  
|  
|  
Db 8 FNV5 12

RESULT 6  
OG6007 PRELIMINARY; PRT; 114 AA.  
AC OG6007;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
GN NDI.  
OS Bufo andrewsi.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
OX NCBL\_TaxID=61428;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RT Tunliye B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal.";  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
DR EMBL; AF004528; AAD05134.1; -;  
DR EMBL; AF004527; AAD05132.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NMDH.DH1.  
DR Pfam; PF00146; NADHdh; 1.  
KW Mitochondrion.  
FT NON\_TER  
SQ SEQUENCE 114 AA; 12906 MW; FB7F83360690E97 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNV5 5  
|  
|  
|  
|  
|  
Db 8 FNV5 12

RESULT 7

OG63322  
ID OG63322 PRELIMINARY; PRT; 114 AA.  
AC OG63322;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
GN NDI.  
OS Bufo exul.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
OX NCBL\_TaxID=30330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RT Tunliye B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal.";  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
DR EMBL; AF004524; AAD05126.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NMDH.DH1.  
DR Pfam; PF00146; NADHdh; 1.  
KW Mitochondrion.  
FT NON\_TER  
SQ SEQUENCE 114 AA; 12886 MW; 7693866FEB1D824 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNV5 5  
|  
|  
|  
|  
|  
Db 8 FNV5 12

RESULT 8  
ID OG63324 PRELIMINARY; PRT; 114 AA.  
AC OG63324;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
GN NDI.  
OS Bufo viridis.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
OX NCBL\_TaxID=30338;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RT Tunliye B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal.";  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
DR EMBL; AF004525; AAD05128.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NMDH.DH1.  
DR Pfam; PF00146; NADHdh; 1.  
KW Mitochondrion.  
FT NON\_TER  
SQ SEQUENCE 114 AA; 12990 MW; B08B6D3AF7E0643 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 DB 8 FNVEY 12

RESULT 9  
 ID 063326 PRELIMINARY; PRT; 114 AA.  
 AC 063326;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
 GN NDI.  
 OS Bufo verricosimus.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
 OX NCBI\_TaxID=61429;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98140324; Pubmed=9479697;  
 RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
 RT "Phylogenetic relationships of toads in the Bufo bufo species group  
 from the eastern escarpment of the Tibetan Plateau: a case of  
 vicariance and dispersal."  
 RT Mol. Phylogenet. Evol. 9:80-87(1998).  
 RL EMBL: AF004526; AAD05130.1; -  
 DR InterPro: IPR001694; Resp-chain\_NADH\_DH1.  
 DR Pfam: PF00146; NADHdh; 1.  
 KW Mitochondrion.  
 FT NON\_TER  
 SQ SEQUENCE 114 AA; 12914 MW; A3799D377A6A27E8 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 DB 8 FNVEY 12

RESULT 10  
 ID 063328 PRELIMINARY; PRT; 114 AA.  
 AC 063328;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
 GN NDI.  
 OS Bufo andrewsi.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
 OX NCBI\_TaxID=61428;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98140324; Pubmed=9479697;  
 RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
 RT "Phylogenetic relationships of toads in the Bufo bufo species group  
 from the eastern escarpment of the Tibetan Plateau: a case of  
 vicariance and dispersal."  
 RT Mol. Phylogenet. Evol. 9:80-87(1998).  
 RL EMBL: AF004529; AAD05136.1; -  
 DR InterPro: IPR001694; Resp-chain\_NADH\_DH1.  
 DR Pfam: PF00146; NADHdh; 1.  
 KW Mitochondrion.

FT NON\_TER 1 1  
 SQ SEQUENCE 114 AA; 12876 MW; E70F9F4F10690E82 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 DB 8 FNVEY 12

RESULT 11  
 ID 063330 PRELIMINARY; PRT; 114 AA.  
 AC 063330;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
 GN NDI.  
 OS Bufo gargarizans.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.  
 OX NCBI\_TaxID=30331;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98140324; Pubmed=9479697;  
 RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
 RT "Phylogenetic relationships of toads in the Bufo bufo species group  
 from the eastern escarpment of the Tibetan Plateau: a case of  
 vicariance and dispersal."  
 RT Mol. Phylogenet. Evol. 9:80-87(1998).  
 RL EMBL: AF004530; AAD05138.1; -  
 DR InterPro: IPR001694; Resp-chain\_NADH\_DH1.  
 DR Pfam: PF00146; NADHdh; 1.  
 KW Mitochondrion.  
 FT NON\_TER  
 SQ SEQUENCE 114 AA; 12918 MW; E3CB884EC1B90E97 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNVEY 5  
 DB 8 FNVEY 12

RESULT 12  
 ID 09MG11 PRELIMINARY; PRT; 118 AA.  
 AC 09MG11;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE NADH DEHYDROGENASE (FRAGMENT).  
 GN NDI.  
 OS Coregonus artedii.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Coregonus.  
 OX NCBI\_TaxID=36181;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Turgeon J., Bernatchez L.;  
 RT "Coregonus artedii phylogeography."  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF246933; AAF90053.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
DR Pfam: PF00146; NADHdh. 1.  
DR PROSITE: PS00668; COMPLEX1\_ND1\_2; 1.  
KW Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 118  
SQ SEQUENCE 118 AA; 12880 MW; 71851F8DA57D4C09 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 118;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 62 FNVEY 66

RESULT 13  
O9MDE1 PRELIMINARY; PRT; 119 AA.  
AC O9MDE1:  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).  
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McCusker M.R., Parkinson E., Taylor E.B.;  
RT "Mitochondrial DNA variation in rainbow trout (Oncorhynchus mykiss)  
RT across its native range: testing biogeographical hypotheses and their  
RT relevance to conservation."  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF254849; AAF71365.1; -;  
DR EMBL: AF254843; AAF71359.1; -;  
DR EMBL: AF254844; AAF71360.1; -;  
DR EMBL: AF254845; AAF71361.1; -;  
DR EMBL: AF254846; AAF71362.1; -;  
DR EMBL: AF254847; AAF71363.1; -;  
DR EMBL: AF254848; AAF71364.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
DR Pfam: PF00146; NADHdh. 1.  
KW Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 119  
SQ SEQUENCE 119 AA; 13174 MW; DED639949FE02174 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 119;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 12 FNVEY 16

RESULT 14  
O9MJM4 PRELIMINARY; PRT; 119 AA.  
AC O9MJM4:  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT 1 (FRAGMENT).  
OS Oncorhynchus clarki (Cutthroat trout) (Salmo clarki).  
OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=30962;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20571467; PubMed=11123621;  
RA McCusker M.R., Parkinson E., Taylor E.B.;  
RT "Mitochondrial DNA variation in rainbow trout (Oncorhynchus mykiss)  
RT across its native range: testing biogeographical hypotheses and their  
RT relevance to conservation."  
RL MOL. ECOL. 9:2089-2108(2000).  
DR EMBL: AF254865; AAF71366.1; -;  
DR InterPro: IPR001694; Resp\_chain\_NADH\_DH1.  
DR Pfam: PF00146; NADHdh. 1.  
KW Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 119  
SQ SEQUENCE 119 AA; 13193 MW; DF6789949EE12074 CRC64;

Query Match 100.0%; Score 28; DB 8; Length 119;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 12 FNVEY 16

RESULT 15  
O958M0 PRELIMINARY; PRT; 121 AA.  
AC O958M0:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT I (FRAGMENT).  
GN ND1.  
OS Rana catesbeiana (Bull frog).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranae;  
OX NCBI\_TaxID=8400;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21184280; PubMed=11286498;  
RA Macey J.R., Strasburg J.L., Brissson J.A., Vredenburg V.T.,  
RA Jennings M., Larson A.;  
RT "Molecular Phylogenetics of Western North American Frogs of the Rana  
RT boylii Species Group."  
RT MOL. Phylogenet. Evol. 19:131-143(2001).  
DR EMBL: AF314016; AAK56866.1; -;  
KW Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 121  
SQ SEQUENCE 121 AA; 13569 MW; D657A0BF338819FF CRC64;

Query Match 100.0%; Score 28; DB 8; Length 121;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNVEY 5  
|||||  
DB 16 FNVEY 20

Search completed: August 20, 2002, 11:31:53  
Job time: 1467 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:15 ; Search time 187.18 Seconds  
(without alignments)  
4.154 Million cell updates/sec

Title: US-09-824-286-14  
Perfect score: 38  
Sequence: 1 KEIHLQ 7

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_032802.\*  
1: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SID55/gcgcdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	19 AAW31651	Human cytokine rec
2	38	100.0	230	15 AAR47151	IL-2 receptor gamm
3	38	100.0	230	16 AAR82934	Interleukin 4 comp
4	38	100.0	232	15 AAR47150	IL-2 receptor gamm
5	38	100.0	347	15 AAR47149	IL-2 receptor gamm
6	38	100.0	369	15 AAR47148	IL-2 receptor gamm
7	38	100.0	482	19 AAW31646	Human cytokine rec
8	38	100.0	691	21 AAY92202	Fusion polypeptide
9	38	100.0	694	21 AAY92201	Fusion polypeptide
10	38	100.0	694	21 AAY92203	Fusion polypeptide
11	33	86.8	1059	21 AAB37570	Wheat starch synth

12	33	86.8	1628	21 AAB37569	Wheat starch synth
13	33	86.8	1628	21 AAB49304	Wheat starch synth
14	32	84.2	233	20 AAY07011	Breast cancer asso
15	32	84.2	233	21 AAB42565	Human OREF ORF239
16	32	84.2	367	20 AAY31735	Human cell cycle r
17	31	81.6	71	21 AAG40993	Zea mays protein f
18	31	81.6	85	21 AAG18933	Zea mays protein f
19	31	81.6	120	21 AAG18932	Zea mays protein f
20	31	81.6	175	21 AAG34058	Zea mays protein f
21	31	81.6	210	21 AAG34057	Zea mays protein f
22	31	81.6	258	21 AAG34056	Zea mays protein f
23	31	81.6	263	22 AAG23964	Novel human diagno
24	31	81.6	677	22 AAB61419	Drosophila melanog
25	31	81.6	756	22 ABB58408	Drosophila melanog
26	31	81.6	785	22 ABB62933	Drosophila melanog
27	30	78.9	231	18 AAM20162	H. pylori inner me
28	30	78.9	431	21 AAB52596	Helicobacter pylor
29	30	78.9	433	17 AAM05196	Helicobacter pylor
30	30	78.9	438	18 AAM20916	H. pylori inner me
31	30	78.9	454	22 AAG92495	C glutamincum prote
32	29	76.3	38	22 AAM96286	Human reproductive
33	29	76.3	38	22 AAM42315	Human breast or ov
34	29	76.3	60	22 AAG99784	ERA binding domain
35	29	76.3	70	22 AAM18654	Peptide #508 enco
36	29	76.3	81	22 AAG40659	Propionibacterium
37	29	76.3	104	21 AAG57029	Arabidopsis thalia
38	29	76.3	107	21 AAG57028	Arabidopsis thalia
39	29	76.3	108	21 AAG57027	Arabidopsis thalia
40	29	76.3	111	22 AAG82931	S. epidermidis ope
41	29	76.3	111	22 AAG83036	S. epidermidis ope
42	29	76.3	133	22 AAM39731	Human polypeptide
43	29	76.3	161	22 AAM41517	Human polypeptide
44	29	76.3	194	21 AAG55688	Arabidopsis thalia
45	29	76.3	197	21 AAG55687	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1	
AAW31651	AAW31651 standard; Peptide; 7 AA.
XX	XX
AC	AAW31651;
XX	XX
DT	21-MAY-1998 (first entry)
XX	XX
DE	Human cytokine receptor gc chain epitope.
XX	XX
KW	Cytokine receptor; gamma common chain; gc chain; human;
KW	blocking agent; monoclonal antibody; CP-88; immunological disease;
KW	myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
KW	insulin-dependent diabetes; inflammatory bowel disease;
KW	sympathetic ophthalmia; uveitis; allergy; asthma; infection;
KW	graft versus host disease; psoriasis; immunosuppressive; therapy;
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO9743416-A1.
XX	XX
PD	20-NOV-1997.
XX	XX
PF	09-MAY-1997; 97WO-US07870.
XX	XX
PR	10-MAY-1996; 96US-0017466.
XX	XX
PA	(BIOJ ) BIOGEN INC.
XX	XX
PI	Benjamin CD, Burkly LC, Hession C, Whitty A;
XX	XX
DR	WPI: 1998-008885/01.
XX	XX

PT Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
PS Claim 24; Page 84; 11pp; English.  
XX  
CC This peptide comprises an epitope of the human cytokine receptor  
CC common gamma (gc) chain (see AAW31646) that is recognised by  
CC gc blocking agents of the invention. 5 Such epitopes (see  
CC AAW31650-54) have been identified. The invention provides  
CC compositions and methods for inhibiting cytokine signalling using  
CC gc chain blocking agents for the treatment of immunological  
CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,  
CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel  
CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,  
CC parasitic infection, graft vs. host disease or psoriasis. A  
CC preferred gc blocking agent is Mab CP.88 or its Fab fragment (see  
CC also AAW31647-48).  
XX  
SQ Sequence 7 AA:  
  
Query Match 100.0%; Score 38; DB 19; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEIHLXQ 7  
DQ 1 keihlyq 7  
|||||  
|  
RESULT 2  
AA047151  
ID AAR47151 standard; Protein: 230 AA.  
XX  
AC AAR47151:  
XX  
XX 13-JUN-1994 (first entry)  
DE  
XX IL-2 receptor gamma chain.  
XX  
XX Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;  
KW rheumatoid arthritis; transplant rejection; primer;  
KW polymerase chain reaction; PCR; amplification.  
XX  
XX Homo sapiens.  
XX  
XX EP578932-A.  
PN  
PD 19-JAN-1994.  
XX  
XX 22-APR-1993; 93EP-0106561.  
PF  
XX 23-APR-1992; 92JP-0104947.  
PR  
XX  
PA (AJIN ) AJINOMOTO KK.  
PA (SUGA/) SUGAMURA K.  
XX  
PI Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;  
PI Suzuki M, Takeshita T;  
XX  
XX WPI: 1994-0117546/03.  
DR N-PSDB; AA054831.  
XX  
XX  
PT DNA and protein sequences of IL-2 gamma chain - useful as immune  
PT regulatory agents for treatment of e.g. rheumatoid arthritis and  
PT transplant rejection  
XX  
PS Disclosure; Page 22-23, 35-36, 50pp; English.  
XX  
XX The human IL-2 receptor gamma chain preform (AAR47148), including the  
CC signal peptide, is encoded by the sequence given in AA054828. The  
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble

CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830.  
CC while a soluble form suitable for expression in prokaryotes (AAR47151)  
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-  
CC terminal sequence of IL-2 receptor gamma chain, and are used to  
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27  
CC are used to obtain the protein given in AAR47151.  
XX  
SQ Sequence 230 AA:  
  
Query Match 100.0%; Score 38; DB 15; Length 230;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEIHLXQ 7  
DQ 98 keihlyq 104  
|||||  
|  
RESULT 3  
AAR82934  
ID AAR82934 standard; Protein: 230 AA.  
XX  
AC AAR82934:  
XX  
XX 26-FEB-1996 (first entry)  
DE  
XX Interleukin 4 component common to the IL-2 receptor gamma chain.  
XX  
XX  
XX Interleukin-4; IL-4; gamma chain component; immunosuppressants;  
KW anti-allergy agent; signal transduction inhibitor; autoimmunity;  
KW disease; anti-inflammatory; anaphylactic shock; bronchial asthma;  
KW Interleukin-2; IL-2; atopic dermatitis; urticaria.  
XX  
XX Homo sapiens.  
OS  
XX JP07149662-A.  
PN  
XX 13-JUN-1995.  
PD  
XX 07-SEP-1994; 94JP-0213706.  
PF  
XX 08-SEP-1993; 93JP-0223574.  
PR  
XX  
PA (AJIN ) AJINOMOTO KK.  
PA (SUGA/) SUGAMURA K.  
XX  
XX WPI: 1995-243601/32.  
DR N-PSDB; AAT04952.  
XX  
XX  
PT Novel interleukin-4 receptor monoclonal antibodies inhibit signal  
PT transmission - useful as immunosuppressants and anti-allergy agents.  
XX  
PS Example 1; Page 9; 11pp; Japanese.  
XX  
XX AAT04952 encodes AAR82934 a component of the IL-4 receptor common to  
CC the IL-2 receptor gamma chain molecule, which was used to generate  
CC anti-IL-4 receptor monoclonal antibodies (mabs). The mabs (IL-4  
CC signal transduction inhibitors) can be used as immunosuppressants  
CC and anti-allergy agents, for the treatment of autoimmune and chronic  
CC inflammatory diseases, e.g. anaphylactic shock, bronchial asthma,  
CC atopic dermatitis and urticaria.  
XX  
SQ Sequence 230 AA:  
  
Query Match 100.0%; Score 38; DB 16; Length 230;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEIHLXQ 7  
DQ 99 keihlyq 105  
|||||  
|

```

RESULT 4
ID AAR47150 standard; Protein; 252 AA.
XX
AC AAR47150;
XX
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KM rheumatoid arthritis; transplant rejection; primer;
XX polymerase chain reaction; PCR; amplification; ss.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..22
FT /label= Sig_Peptide
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
XX (AJIN ) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI; 1994-017546/03.
XX
DR N-PSDB; AA054830.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
XX PS Disclosure; Page 21-22, 34-35; 50pp; English.
XX
XX CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
XX SQ Sequence 252 AA;

```

```

Query Match 100.0%; Score 38; DB 15; Length 252;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 KEIHLXQ 7
Db 120 keihlyq 126

```

```

RESULT 5
ID AAR47149 standard; Protein; 347 AA.
XX
AC AAR47149;
XX

```

```

DT 13-JUN-1994 (first entry)
XX
XX DE IL-2 receptor gamma chain.
XX
XX KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
XX KM rheumatoid arthritis; transplant rejection; primer;
XX KW polymerase chain reaction; PCR; amplification.
XX
XX OS Homo sapiens.
XX
XX PN EP578932-A.
XX
XX PD 19-JAN-1994.
XX
XX PE 22-APR-1993; 93EP-0106561.
XX
XX PR 23-APR-1992; 92JP-0104947.
XX
XX (AJIN ) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
XX DR WPI; 1994-017546/03.
XX
XX DR N-PSDB; AA054829.
XX
XX PT DNA and protein sequences of IL-2 gamma chain - useful as immune
XX PT regulatory agents for treatment of e.g. rheumatoid arthritis and
XX PT transplant rejection
XX
XX PS Claim 4; Page 41; 50pp; English.
XX
XX CC The human IL-2 receptor gamma chain preform (AAR47148), including the
XX CC signal peptide, is encoded by the sequence given in AA054828. The
XX CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
XX CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
XX CC while a soluble form suitable for expression in prokaryotes (AAR47151)
XX CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
XX CC terminal sequence of IL-2 receptor gamma chain, and are used to
XX CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
XX CC are used to obtain the protein given in AAR47151.
XX
XX SQ Sequence 347 AA;

```

```

Query Match 100.0%; Score 38; DB 15; Length 347;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 KEIHLXQ 7
Db 98 keihlyq 104

```

```

RESULT 6
ID AAR47148 standard; Protein; 369 AA.
XX
AC AAR47148;
XX

```

```

DT 13-JUN-1994 (first entry)
XX

```

```

DE IL-2 receptor gamma chain.
XX

```

```

KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KM rheumatoid arthritis; transplant rejection; primer; PCR;
XX polymerase chain reaction; amplification; ss.
XX
XX OS Homo sapiens.
XX
XX PN EP578932-A.
XX
XX PD 19-JAN-1994.
XX
XX PE 22-APR-1993; 93EP-0106561.
XX
XX PR 23-APR-1992; 92JP-0104947.
XX
XX (AJIN ) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
XX DR WPI; 1994-017546/03.
XX
XX DR N-PSDB; AA054829.
XX
XX PT DNA and protein sequences of IL-2 gamma chain - useful as immune
XX PT regulatory agents for treatment of e.g. rheumatoid arthritis and
XX PT transplant rejection
XX
XX PS Claim 4; Page 41; 50pp; English.
XX
XX CC The human IL-2 receptor gamma chain preform (AAR47148), including the
XX CC signal peptide, is encoded by the sequence given in AA054828. The
XX CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
XX CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
XX CC while a soluble form suitable for expression in prokaryotes (AAR47151)
XX CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
XX CC terminal sequence of IL-2 receptor gamma chain, and are used to
XX CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
XX CC are used to obtain the protein given in AAR47151.
XX
XX SQ Sequence 347 AA;

```

```

Key Location/Qualifiers
FH Peptide 1..22
FT

```

```

FT  /label= Sig-peptide
XX
XX  EP578932-A.
XX
XX  19-JAN-1994.
XX
XX
XX  22-APR-1993; 93EP-0106561.
XX
XX  23-APR-1992; 92JP-0104947.
XX
XX  (AJIN ) AJINOMOTO KK.
XX  (SUGA/) SUGAMURA K.
XX
XX  Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
XX  Suzuki M, Takeshita T;
XX  WPI: 1994-017546/03.
XX  N-PSDB: AAQ54828.
XX
XX  DNA and protein sequences of IL-2 gamma chain - useful as immune
XX  regulatory agents for treatment of e.g. rheumatoid arthritis and
XX  transplant rejection
XX
XX  Disclosure: Page 16-17, 29-30; 50pp; English.
XX
XX  The human IL-2 receptor gamma chain preform (AAK47148), including the
XX  signal peptide, is encoded by the sequence given in AAQ54828. The
XX  mature protein (AAK47149) is encoded by sequence AAQ54829. A soluble
XX  form of IL-2 receptor gamma chain (AAK47150) is encoded by AAQ54830,
XX  while a soluble form suitable for expression in prokaryotes (AAK47151)
XX  is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-
XX  terminal sequence of IL-2 receptor gamma chain, and are used to
XX  isolate IL2 receptor gamma chain receptor cDNA. Primers AAQ54826-27
XX  are used to obtain the protein given in AAK47151.
XX
XX  Sequence 369 AA;
SQ
Query Match 100.0%; Score 38; DB 15; Length 369;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHYQ 7
DB 120 keihlyq 126
RESULT 7
AAW31646
ID AAW31646 standard; Protein: 482 AA.
XX
XX  AAW31646;
XX
XX  21-MAY-1998 (first entry)
XX
XX  Human cytokine receptor gc chain-Ig fusion protein.
XX
XX  Cytokine receptor; gamma common chain; gc chain; human;
XX  blocking agent; monoclonal antibody; CP.B8; Immunological disease;
XX  myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
XX  insulin-dependent diabetes; inflammatory bowel disease;
XX  sympathetic ophthalmia; uveitis; allergy; asthma; infection;
XX  graft versus host disease; psoriasis; immunosuppressive; therapy.
XX
XX  Chimeric - Homo sapiens.
XX
XX
XX  Key Location/Qualifiers
XX  Protein 1..254
XX  Protein /note= "gc chain N-terminal region"
XX  Protein 235..482
XX  Region /note= "IgG1 constant region"
XX  Region 255..264
XX  Region /note= "IgG1 hinge region"

```

```

FT  Domain 264..482
FT  /note= "IgG1 CH2 and CH3 constant domains1"
XX
XX  WO9743416-A1.
XX
XX  20-NOV-1997.
XX
XX  09-MAY-1997; 97WO-US07870.
XX
XX  10-MAY-1996; 96US-0017466.
XX
XX  (BIOJ ) BIOGEN INC.
XX
XX  Benjamin CD, Burkly LC, Hession C, Whitty A;
XX  WPI: 1998-008885/01.
XX  N-PSDB: AAT97439.
XX
XX  Blocking agents of the gamma common chain of cytokine receptors -
XX  particularly monoclonal antibodies, used to induce T cell anergy for
XX  treatment of immunological diseases
XX
XX  Example 1: Page 79-80; 111pp; English.
XX
XX  This polypeptide comprises a fusion between the N-terminal 254
XX  amino acids of the human mature cytokine receptor gamma common (gc)
XX  chain and the hinge region and CH2 and CH3 constant domains of
XX  human IgG1. The fusion was expressed from clone pLB001 (see
XX  AAT97439) in COS-7 cells, and used to generate murine anti-human gc
XX  specific monoclonal antibodies (MAbs), including CP.B8 produced by
XX  hybridoma ATCC HB 12107. The invention provides compositions and
XX  methods for inhibiting cytokine signalling using gc chain blocking
XX  agents for the treatment of immunological diseases such as
XX  myasthenia gravis, rheumatoid arthritis, lupus, multiple sclerosis,
XX  insulin-dependent diabetes, inflammatory bowel disease, sympathetic
XX  ophthalmia, uveitis, allergy, asthma, parasitic infection, graft
XX  vs. host disease or psoriasis. A preferred gc blocking agent is
XX  Mab CP.B8 or its Fab fragment (see also AAW31647-48).
XX
XX  Sequence 482 AA;
SQ
Query Match 100.0%; Score 38; DB 19; Length 482;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHYQ 7
DB 120 keihlyq 126
RESULT 8
AAV92202
ID AAV92202 standard; Protein: 691 AA.
XX
XX  AAV92202;
XX
XX  01-AUG-2000 (first entry)
XX
XX  Fusion polypeptide 603, IL-4 trap.
XX
XX  IL-4 trap; cytokine; antagonist; CNMF; receptor; fusion protein;
XX  cyostatic; immunomodulator; osteopathic.
XX
XX  Synthetic.
XX  Homo sapiens.
XX
XX  WO200018932-A2.
XX  06-APR-2000.
XX  22-SEP-1999; 99WO-US22045.

```



```
PR 25-SEP-1998; 980S-0101858.
PR 19-MAY-1999; 990S-0313942.
XX
PA (REGE-) REGENERON PHARM INC.
XX
PI Stahl N, Yancopoulos GD;
XX
DR WPI; 2000-293165/25.
DR N-PSDB; AAA09044.
XX
PT Isolated nucleic acid molecule for treating cytokine-related diseases
PT or disorders encodes a fusion polypeptide capable of binding a cytokine
PT to form a nonfunctional complex
XX
PS Example 6; Fig 22A-D; 152pp; English.
XX
CC This sequence shows fusion polypeptide 603, which is capable of
CC binding cytokine IL-4 to form a non-functional complex.
CC The invention concerns production of antagonists to any cytokine that
CC utilizes an alpha specificity determining component, which when combined
CC with the cytokine, binds to a first beta signal transducing component to
CC form a non-functional intermediate which then binds to a second beta
CC signal transducing component causing beta-receptor dimerization, the
CC soluble alpha specificity determining component of the receptor
CC (SR-alpha) and the extracellular domain of the first beta signal
CC transducing component of the cytokine receptor (beta-1) are combined to
CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
CC cytokine by binding the cytokine to form a non-functional complex. The
CC receptor components are shared by cytokines such as the CNTF (ciliary
CC neurotrophic factor) family of cytokines. The invention provides the
CC basis for the development of IL-6 antagonists, as they show that if, in
CC the presence of a ligand, a non-functional intermediate complex,
CC consisting of the ligand, its alpha receptor and its beta-1 receptor
CC component, can be formed, it will effectively block the action of the
CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
CC of the extracellular domains of the alpha specificity determining
CC components of their receptors and the extracellular domain of gp130.
CC The resultant heterodimers, function as high-affinity traps, rendering
CC the cytokine inaccessible to form a signal transducing complex with the
CC native membrane-bound forms of their receptor. The nucleic acids and
CC polypeptides are useful for treating cytokine-related diseases or
CC disorders such as osteoporosis and primary and secondary effects of
CC cancer including multiple myeloma or cachexia.
CC
XX
SQ Sequence 691 AA;
XX
Query Match 100.0%; Score 38; DB 21; Length 691;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLQ 7
DB 120 keihlyg 126
XX
RESULT 9
AA92201
ID AAY92201 standard; Protein; 694 AA.
XX
AC AAY92201;
XX
DT 01-AUG-2000 (first entry)
XX
DE Fusion polypeptide 424, IL-4 trap.
XX
KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
XX
KW cytosolic; immunomodulator; osteopathic.
XX
OS Synthetic.
XX
OS Homo sapiens.
XX
PN WO200018932-A2.
```

```
XX
XX 06-APR-2000.
PD
XX
XX 22-SEP-1999; 99WO-US22045.
PE
XX
XX 25-SEP-1998; 980S-0101858.
PR 19-MAY-1999; 990S-0313942.
XX
XX (REGE-) REGENERON PHARM INC.
PA
XX
PI Stahl N, Yancopoulos GD;
XX
DR WPI; 2000-293165/25.
DR N-PSDB; AAA09043.
XX
PT Isolated nucleic acid molecule for treating cytokine-related diseases
PT or disorders encodes a fusion polypeptide capable of binding a cytokine
PT to form a nonfunctional complex
XX
PS Example 6; Fig 21; 152pp; English.
XX
CC This sequence shows fusion polypeptide 424, which is capable of
CC binding cytokine IL-4 to form a non-functional complex.
CC The invention concerns production of antagonists to any cytokine that
CC utilizes an alpha specificity determining component, which when combined
CC with the cytokine, binds to a first beta signal transducing component to
CC form a non-functional intermediate which then binds to a second beta
CC signal transducing component causing beta-receptor dimerization, the
CC soluble alpha specificity determining component of the receptor
CC (SR-alpha) and the extracellular domain of the first beta signal
CC transducing component of the cytokine receptor (beta-1) are combined to
CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the
CC cytokine by binding the cytokine to form a non-functional complex. The
CC receptor components are shared by cytokines such as the CNTF (ciliary
CC neurotrophic factor) family of cytokines. The invention provides the
CC basis for the development of IL-6 antagonists, as they show that if, in
CC the presence of a ligand, a non-functional intermediate complex,
CC consisting of the ligand, its alpha receptor and its beta-1 receptor
CC component, can be formed, it will effectively block the action of the
CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers
CC of the extracellular domains of the alpha specificity determining
CC components of their receptors and the extracellular domain of gp130.
CC The resultant heterodimers, function as high-affinity traps, rendering
CC the cytokine inaccessible to form a signal transducing complex with the
CC native membrane-bound forms of their receptor. The nucleic acids and
CC polypeptides are useful for treating cytokine-related diseases or
CC disorders such as osteoporosis and primary and secondary effects of
CC cancer including multiple myeloma or cachexia.
CC
XX
SQ Sequence 694 AA;
XX
Query Match 100.0%; Score 38; DB 21; Length 694;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEIHLQ 7
DB 120 keihlyg 126
XX
RESULT 10
AA92203
ID AAY92203 standard; Protein; 694 AA.
XX
AC AAY92203;
XX
DT 01-AUG-2000 (first entry)
XX
DE Fusion polypeptide 622, IL-4 trap.
XX
KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;
XX
KW cytosolic; immunomodulator; osteopathic.
```

XX OS Synthetic.  
 XX OS Homo sapiens.  
 XX PN WO200018932-A2.  
 XX PD 06-APR-2000.  
 XX PF 22-SEP-1999; 99WO-US22045.  
 XX PR 25-SEP-1998; 98US-0101858.  
 XX PR 19-MAY-1999; 99US-0313942.  
 XX PA (REGG-) REGENERON PHARM INC.  
 XX PI Stahl N, Yancopoulos GD;  
 XX DR WPI: 2000-293165/25.  
 XX DR N-PSDB; AAA09045.  
 XX PS Isolated nucleic acid molecule for treating cytokine-related diseases or disorders encodes a fusion polypeptide capable of binding a cytokine to form a nonfunctional complex

XX PS Example 6; Fig 23A-D; 152pp; English.

CC This sequence shows fusion polypeptide 622, which is capable of binding cytokine IL-4 to form a non-functional complex.  
 CC The invention concerns production of antagonists to any cytokine that utilizes an alpha specificity determining component, which when combined with the cytokine, binds to a first beta signal transducing component to form a non-functional intermediate which then binds to a second beta signal transducing component causing beta-receptor dimerization, the soluble alpha specificity determining component of the receptor (SR-alpha) and the extracellular domain of the first beta signal transducing component of the cytokine receptor (beta-1) are combined to form heterodimers (SR-alpha:beta-1) that act as antagonist to the cytokine by binding the cytokine to form a non-functional complex. The receptor components are shared by cytokines such as the CNTF (ciliary neurotrophic factor) family of cytokines. The invention provides the basis for the development of IL-6 antagonists, as they show that if, in the presence of a ligand, a non-functional intermediate complex, consisting of the ligand, its alpha receptor and its beta-1 receptor component, can be formed, it will effectively block the action of the ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers of the extracellular domains of the alpha specificity determining components of their receptors and the extracellular domain of gp130.  
 CC The resultant heterodimers, function as high-affinity traps, rendering the cytokine inaccessible to form a signal transducing complex with the native membrane-bound forms of their receptor. The nucleic acids and polypeptides are useful for treating cytokine-related diseases or disorders such as osteoporosis and primary and secondary effects of cancer including multiple myeloma or cachexia.

CC Sequence 694 AA;  
 XX SQ

Query Match 100.0%; Score 38; DB 21; Length 694;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLQY 7  
 I:|||||  
 Db 120 kehllyq 126

RESULT 11  
 AAB37570  
 ID AAB37570 standard; Protein; 1059 AA.  
 AC AAB37570;  
 XX  
 DT 01-MAR-2001 (first entry)

XX DE Wheat starch synthase III SEQ ID NO: 10.  
 XX KW wheat; starch synthase; SSII; SSIII; starch content; starch synthesis;  
 XX KW food product; adhesive.  
 XX OS Trilicium aestivum.  
 XX PN WO200066745-A1.  
 XX PD 09-NOV-2000.  
 XX PF 28-APR-2000; 2000WO-AU00385.  
 XX PR 29-APR-1999; 99AU-0000052.  
 XX PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 XX PA (GOOD-) GOODMAN FIELDER LTD.  
 XX PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.  
 XX PI Morell M, Li Z, Rahman S, Appels R;  
 XX DR WPI: 2000-647602/62.  
 XX DR N-PSDB; AAC86414.  
 XX PS Nucleic acid molecules encoding wheat starch synthase (WST-I and WST-II, useful in modifying plant starch content and/or composition - Claim 9; Page 183-186; 211pp; English.

CC The present invention relates to novel protein and coding sequences from wheat. The proteins are wheat starch synthases, designated SSII and SSIII. These can be used in the modification of plant starch content or composition, and to screen plants to identify mutations which affect starch content and composition. The starch can then be used in food products, such as flour, and in films, coatings, adhesives, building materials and packaging materials.

CC Sequence 1059 AA;  
 XX SQ

Query Match 86.8%; Score 33; DB 21; Length 1059;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLQY 7  
 I:|||||  
 Db 663 kdhllyq 669

RESULT 12  
 AAB37569  
 ID AAB37569 standard; Protein; 1628 AA.  
 AC AAB37569;  
 XX  
 DT 01-MAR-2001 (first entry)  
 XX  
 DE Wheat starch synthase III SEQ ID NO: 8.  
 XX wheat; starch synthase; SSII; SSIII; starch content; starch synthesis;  
 XX KW food product; adhesive.  
 XX OS Trilicium aestivum.  
 XX PN WO200066745-A1.  
 XX PD 09-NOV-2000.  
 XX PF 28-APR-2000; 2000WO-AU00385.  
 XX PR 29-APR-1999; 99AU-0000052.

PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
 PA (GOOD-) GOODMAN FIELDER LTD.  
 PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.  
 XX  
 PI Morell M, Li Z, Rahman S, Appels R;  
 XX  
 DR WPI; 2000-647602/62.  
 DR N-PSDB; AAC86413.  
 XX  
 PT Nucleic acid molecules encoding wheat starch synthase (WST)-I and  
 PT WST-II, useful in modifying plant starch content and/or composition -  
 XX  
 PS Claim 9; Page 172-179; 211pp; English.  
 XX  
 CC The present invention relates to novel protein and coding sequences from  
 CC wheat. The proteins are wheat starch synthases, designated SSII and  
 CC SSIII. These can be used in the modification of plant starch content or  
 CC composition, and to screen plants to identify mutations which affect  
 CC starch content and composition. The starch can then be used in food  
 CC products, such as flour, and in films, coatings, adhesives, building  
 CC materials and packaging materials.  
 XX  
 SQ Sequence 1628 AA;

Query Match 86.8%; Score 33; DB 21; Length 1628;  
 Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLVQ 7  
 I::IIII  
 Db 1232 kdhllyg 1238

## RESULT 13

AAB49304 standard; Protein; 1628 AA.

XX AAB49304;

DT 01-MAR-2001 (first entry)

DE Wheat starch synthase III protein.

KW Wheat; starch synthase; SSII; SSIII; starch content; starch synthesis;

KM food product; adhesive.

XX Triticum aestivum.

XX WO200066745-A1.

PD 09-NOV-2000.

PF 28-APR-2000; 2000WO-AU00385.

PR 29-APR-1999; 99AU-0000052.

PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.

PA (GOOD-) GOODMAN FIELDER LTD.

PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.

PI Morell M, Li Z, Rahman S, Appels R;

DR WPI; 2000-647602/62.

PT Nucleic acid molecules encoding wheat starch synthase (WST)-I and  
 PT WST-II, useful in modifying plant starch content and/or composition -  
 XX  
 PS Example 13; Fig 7; 211pp; English.

CC The present invention relates to novel protein and coding sequences from  
 CC wheat. The proteins are wheat starch synthases, designated SSII and  
 CC SSIII. These can be used in the modification of plant starch content or

CC composition, and to screen plants to identify mutations which affect  
 CC starch content and composition. The starch can then be used in food  
 CC products, such as flour, and in films, coatings, adhesives, building  
 CC materials and packaging materials.

SQ Sequence 1628 AA;

Query Match 86.8%; Score 33; DB 21; Length 1628;  
 Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLVQ 7  
 I::IIII  
 Db 1232 kdhllyg 1238

## RESULT 14

AAY07011 standard; Protein; 164 AA.

XX AAY07011;

DT 02-JUL-1999 (first entry)

DE Breast cancer associated antigen precursor sequence.

KW Cancer associated antigen; diagnosis; research; treatment; human;

KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer.

XX Homo sapiens.

XX WO9904265-A2.

PD 28-JAN-1999.

PF 15-JUL-1998; 98WO-US14679.

PR 22-JUN-1998; 98US-0102322.

PR 17-JUL-1997; 97US-0896164.

PR 10-OCT-1997; 97US-0061599.

PR 10-OCT-1997; 97US-0061765.

PR 10-OCT-1997; 97US-0948705.

PR 11-OCT-1997; 97GB-0021697.

XX (LUDM-) LUDMIG INST CANCER RES.

PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;

PI Freundschuh M, Sahin U, Scanlan MJ, Stockert E;  
 PI Tureci O;

DR WPI; 1999-132448/11.

PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure; Page 393-394; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

CC Lung cancer.  
 XX Sequence 164 AA;

Query Match 84.2%: Score 32; DB 20; Length 164;  
 Best Local Similarity 71.4%: Pred. NO. 58;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLQ 7  
 :||:|  
 Db 99 eeihlyq 105

RESULT 15  
 AAB42565  
 ID AAB42565 standard; Protein: 233 AA.

AC AAB42565;

DF 08-FEB-2001 (first entry)

DE Human ORFX ORF2329 polypeptide sequence SEQ ID NO:4658.

XX  
 KW Human: open reading frame: ORFX: detection: cytostatic; hepatotropic;  
 KW vulnary; antiposiatric; antiparkinsonian; neutropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

OS WO200058473-A2.

PN 05-OCT-2000.

PD 31-MAR-2000; 2000WO-US08621.

PE 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shmkets RA, Leach M;

XX WPI: 2000-602362/57.

DR N-PSDB: AAC6774.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 3855-3856; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antiposiatric; antiparkinsonian; neutropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; antidiabetic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antineoplastic; antibacterial; antiviral; antifungal; antineumatic;

CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 233 AA;

Query Match 84.2%: Score 32; DB 21; Length 233;  
 Best Local Similarity 71.4%: Pred. NO. 85;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLQ 7  
 :||:|  
 Db 193 eeihlyq 199

Search completed: August 20, 2002, 11:07:16  
 Job time: 6101 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:34 ; Search time 100.84 Seconds  
(without alignments)  
6.670 Million cell updates/sec

Title: US-09-824-286-14  
Perfect score: 38  
Sequence: 1 KEIHLYQ 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	38	100.0	369	2 A42565	interleukin-2 rece
2	32	84.2	257	2 J00403	hypothetical 28.1k
3	32	84.2	312	4 A40473	RNA-directed RNA p
4	32	84.2	433	2 A10536	probable membrane
5	32	84.2	551	2 T05667	probable growth re
6	31	81.6	233	2 S75267	hypothetical prote
7	31	81.6	294	2 H82272	conserved hypothet
8	31	81.6	373	2 A55718	interleukin-2 rece
9	31	81.6	472	2 S28026	nuclear pore compl
10	31	81.6	594	2 T18628	hypothetical prote
11	31	81.6	673	2 H69828	ABC transporter (A
12	30	78.9	135	2 B95172	hypothetical prote
13	30	78.9	135	2 B98038	hypothetical prote
14	30	78.9	188	2 S74963	hypothetical prote
15	30	78.9	213	2 E83420	endopeptidase Clp
16	30	78.9	215	1 VCMGPM	coat protein - pap
17	30	78.9	223	2 F86728	beta-subunit of L-
18	30	78.9	319	2 D90482	dehydrogenase, pro
19	30	78.9	330	2 AB3267	hypothetical prote
20	30	78.9	357	2 S21209	4-hydroxyphenylpyr
21	30	78.9	433	2 G64594	hemolysin secretio
22	30	78.9	433	2 H71917	methylin-accepting c
23	30	78.9	456	2 T40416	hypothetical prote
24	30	78.9	940	2 H82159	probable helicase
25	30	78.9	1322	2 H86196	hypothetical prote
26	30	78.9	4196	2 T43274	dynein heavy chain
27	29	76.3	129	2 B82745	hypothetical prote
28	29	76.3	146	2 T15533	hypothetical prote
29	29	76.3	153	2 D64628	hypothetical prote

30	29	76.3	153	2 E71888	hypothetical prote
31	29	76.3	194	2 C96740	hypothetical prote
32	29	76.3	197	2 T33661	hypothetical prote
33	29	76.3	229	2 AC1110	conserved hypothet
34	29	76.3	259	2 AG1471	class I histocompa
35	29	76.3	328	2 S39599	hypothetical prote
36	29	76.3	347	2 T32768	hypothetical prote
37	29	76.3	357	2 T32881	CMP-NeuN-GM3 alp
38	29	76.3	380	2 JC6321	alpha-2,8-sialytra
39	29	76.3	380	2 A56950	carboxynorspermidl
40	29	76.3	387	2 S77268	probable serine ca
41	29	76.3	425	2 E84631	hypothetical prote
42	29	76.3	484	2 A12527	ATP-dependent DNA
43	29	76.3	491	2 G83850	translation releas
44	29	76.3	548	2 C82839	hypothetical prote
45	29	76.3	713	2 H64464	hypothetical prote

## ALIGNMENTS

## RESULT 1

A42565  
interleukin-2 receptor gamma chain - human  
C:Species: Homo sapiens (man)  
C:Date: 04-Mar-1993 #sequence.revision 18-Nov-1994 #text\_change 20-Jun-2000  
C:Accession: A42565; A46591; I54332  
R:Takeshita, T.; Asao, H.; Ohtani, K.; Ishii, N.; Kumaki, S.; Tanaka, N.; Munakata, I  
Science 257, 379-382, 1992  
A:Title: Cloning of the gamma chain of the human IL-2 receptor.  
A:Reference number: A42565; MUID:92335883  
A:Accession: A42565  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid: protein  
A:Residues: 1-369 <TRK>  
A:Cross-references: GB:011086; NID:g303611; PIDN:BA01857.1; PID:g219690  
A:Experimental source: MOLT beta lymphoid cells  
A:Note: sequence extracted from NCBI backbone (NCBIP:109167)  
R:Noguchi, M.; Adelstein, S.; Cao, X.; Leonard, W.J.  
J. Biol. Chem. 268, 13601-13608, 1993  
A:Title: Characterization of the human interleukin-2 receptor gamma chain gene.  
A:Reference number: A46591; MUID:93293887  
A:Accession: A46591  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-369 <RNS>  
A:Cross-references: GB:012183; NID:g307056; PIDN:AA59145.1; PID:g307058  
R:Puck, J.M.; Deschenes, S.M.; Porter, J.C.; Dutra, A.S.; Brown, C.J.; Willard, H.F.,  
Hum. Mol. Genet. 2, 1099-1104, 1993  
A:Title: The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-1;  
A:Reference number: I54332; MUID:94004847  
A:Accession: I54332  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-369 <RE2>  
A:Cross-references: GB:019546; NID:g349631; PIDN:AAC37524.1; PID:g349632  
C:Genetics:  
A:Gene: GDB:112RG; SCIDX1; IMD4  
A:Cross-references: GDB:134807; OMIM:308380  
A:Map position: Xq13.1-Xq13.1  
A:Intons: 39/1; 90/2; 152/1; 198/3; 253/1; 285/2; 308/3  
A:Note: defects are associated with an X-linked form of severe combined immunodeficit  
C:Superfamily: interleukin-2 receptor gamma chain  
C:Keywords: cytokine receptor; duplication; immunodeficiency; severe combined immunon

## Query Match

100.0%; Score 38; DB 2; Length 369;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLYQ 7

DB 120 KEIHLYQ 126

RESULT 2  
J00403  
hypothetical 28.1k protein - clover yellow mosaic virus  
N:Contains: coat protein  
C:Species: clover yellow mosaic virus  
C>Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 20-Jun-2000  
R:Abouhaidar, M.G.; Lai, R.  
J. Gen. Virol. 70, 1871-1875, 1989  
A:Title: Nucleotide sequence of the 3'-terminal region of clover yellow mosaic virus RNA  
A:Reference number: J00402; MUID:89293092  
A:Accession: J00403  
A:Molecule type: genomic RNA  
A:Residues: 1-257 <ABO>  
A:Cross-references: GB:D00465; NID:g221218; PIDN:BAA00373.1; PID:g221220  
A>Note: It is possible that the coat protein of CYMV is first translated as a 28.1k prot  
F:46-257/Product: coat protein #status predicted <COP>

Query Match 84.2%; Score 32; DB 2; Length 257;  
Best Local Similarity 71.4%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7  
|:||||  
Db 216 KNVHLXQ 222

RESULT 3  
A40473  
RNA-directed RNA polymerase (EC 2.7.7.48) / coat protein mutant fusion protein - clover  
N:Contains: coat protein; RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: clover yellow mosaic virus  
C>Date: 29-Jan-1993 #sequence\_revision 24-Apr-1997 #text\_change 19-May-2000  
R:White, K.A.; Bancroft, J.B.; Mackie, G.A.  
Virology 183, 479-486, 1991  
A:Title: Defective RNAs of clover yellow mosaic virus encode nonstructural/coat protein  
A:Reference number: A40473; MUID:91306431  
A:Accession: A40473  
A:Molecule type: genomic RNA  
A:Residues: 1-312 <WHI>  
A:Cross-references: GB:M63511; NID:g323436; PIDN:AAA42935.1; PID:g323437  
A>Note: fragments of three other mutant fusion proteins are presented  
C:Keywords: coat protein; fusion protein; nucleotidyltransferase  
F:1-221/Region: RNA-directed RNA polymerase  
F:222-312/Region: coat protein

Query Match 84.2%; Score 32; DB 4; Length 312;  
Best Local Similarity 71.4%; Pred. No. 33;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7  
|:||||  
Db 271 KNVHLXQ 277

RESULT 4  
A10536  
Probable membrane protein STY0305 [imported] - Salmonella enterica subsp. enterica serov  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A>Note: This species has also been called Salmonella typhi  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A:Reference number: AB0502; PMID:11677608  
A:Accession: A10536  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08737.1; PID:g16501559; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY0305

Query Match 84.2%; Score 32; DB 2; Length 433;  
Best Local Similarity 71.4%; Pred. No. 46;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7  
|:||||  
Db 96 REMHLXQ 102

RESULT 5  
T05667  
probable growth regulator F22113.160 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
R:Bevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer,  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: T05667  
A:Accession: T05667  
A:Molecule type: DNA  
A:Residues: 1-551 <BEV>  
A:Cross-references: EMBL:AL035539  
A:Experimental source: cultivar Columbia; BAC clone F22113  
C:Genetics:  
A:Map position: 4  
A:Introns: 85/3; 116/1; 141/3; 171/2; 234/3; 295/2; 329/3; 440/2  
A>Note: F22113.160

Query Match 84.2%; Score 32; DB 2; Length 551;  
Best Local Similarity 83.3%; Pred. No. 59;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 6  
|:||||  
Db 360 KEVHLX 365

RESULT 6  
S75267  
hypothetical protein slr1039 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O.; Kanehisa, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
S.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S75267  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <KAN>  
A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAA17181.1; PID:d101  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Superfamily: bioc homology  
F:43-148/Domain: bioc homology <BIOC>

Query Match 81.6%; Score 31; DB 2; Length 233;

Best Local Similarity 83.3%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 EIHLYQ 7  
|:|||||

Db 208 EIHLYQ 213

RESULT 7

H82272

Conserved hypothetical protein VC0853 [imported] - *Vibrio cholerae* (strain N16961 serogr C:Species: *Vibrio cholerae*

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: H82272

R:Heidelberg, J.F.; Elsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: A82035; MUID:20406833

A:Accession: H82272

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-294 <HEI>

A:Cross-references: GB:AE004170; GB:AE003852; NID:g9655298; PIDN:AAF94015.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0853

C:Superfamily: conserved hypothetical protein HI0072

Query Match 81.6%; Score 31; DB 2; Length 294;  
Best Local Similarity 83.3%; Pred. No. 50;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 EIHLYQ 7  
|:|||||

Db 259 EIHLYQ 264

RESULT 8

A55718

Interleukin-2 receptor gamma chain precursor - dog

C:Species: *Canis lupus familiaris* (dog)

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 23-Jul-1999

C:Accession: A55718

R:Henhorn, P.S.; Somberg, R.L.; Fintiani, V.M.; Puck, J.M.; Patterson, D.F.; Felsburg, F.

Genomics 23, 69-74, 1994

A:Title: IL-2Rgamma gene microdeletion demonstrates that canine X-linked severe combined

A:Reference number: A55718; MUID:95130114

A:Accession: A55718

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-373 <HEN>

A:Cross-references: GB:U04361; NID:9517411; PIDN:AAC48403.1; PID:9517412

C:Superfamily: interleukin-2 receptor gamma chain

C:Keywords: cytokine receptor; duplication

Query Match 81.6%; Score 31; DB 2; Length 373;  
Best Local Similarity 71.4%; Pred. No. 64;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLYQ 7  
|:|||||

Db 120 KEIHLYQ 126

RESULT 9

S28026

nuclear pore complex protein NUP49 - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: nuclear pore complex protein NSP49; nucleoporin NSP49; protein G1648;

C:Species: *Saccharomyces cerevisiae*

C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000

C:Accession: S28026; C44402; S52458; S59239; S64189; S28538

R:Wimmer, C.; Doye, V.; Grandi, P.; Neubass, U.; Hurt, E.C.

EMBO J. 11, 5051-5061, 1992

A:Title: A new subclass of nucleoporins that functionally interact with nuclear pore

A:Reference number: S28026; MUID:93099880

A:Accession: S28026

A:Molecule type: DNA

A:Residues: 1-472 <NIM>

A:Cross-references: EMBL:X68109; NID:g4055; PIDN:CAA8229.1; PID:g4056

R:Wente, S.R.; Rout, M.P.; Blobel, G.

J. Cell Biol. 119, 705-723, 1992

A:Title: A new family of yeast nuclear pore complex proteins.

A:Reference number: A44402; MUID:93054906

A:Accession: C44402

A:Molecule type: DNA

A:Residues: 1-472 <MEN>

A:Cross-references: EMBL:Z15040; NID:g4077; PIDN:CAV78758.1; PID:g4078

A:Note: sequence extracted from NCBI backbone (NCBI:P:117140)

R:Bertrani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.

submitted to the EMBL data library, February 1995

A:Description: The sequence of 11.1kb fragment on the left arm to *Saccharomyces cere*

A:Reference number: S52454

A:Accession: S52458

A:Molecule type: DNA

A:Residues: 1-472 <BER>

A:Cross-references: EMBL:X84705; NID:g677853; PID:g677857

R:Bertrani, I.; Coglievina, M.; Zaccaria, P.; Klima, R.; Bruschi, C.V.

Yeast 11, 1187-1194, 1995

A:Title: The sequence of an 11.1 kb fragment on the left arm of *Saccharomyces cerevi*

A:Reference number: S59235; MUID:96109931

A:Accession: S59239

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-472 <BER>

A:Cross-references: EMBL:X84705; NID:g677853; PID:CAA59181.1; PID:g677857

R:Bruschi, C.V.; Coglievina, M.; Bertrani, I.; Klima, R.; Zaccaria, P.; Delneri, D.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64183

A:Accession: S64189

A:Molecule type: DNA

A:Residues: 1-472 <BRU>

A:Cross-references: EMBL:Z72694; NID:g1322775; PID:g1322776; MIPS:YGL172w

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:NUP49; NSP49

A:Cross-references: SGD:S0003140; MIPS:YGL172w

A:Map position: 7L

C:Keywords: nucleus

Query Match 81.6%; Score 31; DB 2; Length 472;  
Best Local Similarity 71.4%; Pred. No. 82;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLYQ 7  
|:|||||

Db 375 KEIHLYQ 381

RESULT 10

T18628

hypothetical protein B0001.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18628

R:Sims, M.

submitted to the EMBL data library, February 1996

A:Reference number: Z18999

A:Accession: T18628

A:Status: preliminary; translated from GB/EMBL/DBU

A:Residues: 1-594 <NLR>  
A:Cross-references: EMBL:269634; PIDN:CAA93452.1; GSPDB:GN00022; CESP:B0001.3  
A:Experimental source: clone B0001  
A:Genetics:  
A:Gene: CESP:B0001.3  
A:Map position: 4  
A:Introns: 9/2; 20/3; 69/2; 109/3; 154/3; 226/2; 340/3; 493/3; 555/3

Query Match  
Best Local Similarity 81.6%; Score 31; DB 2; Length 594;  
57.1%; Pred. No. 1e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7  
||:||||  
D5 217 KDVLHYE 223

## RESULT 11

H69828

ABC transporter (ATP-binding protein) homolog yheH - Bacillus subtilis  
C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 02-Feb-2001  
C:Accession: H69828

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Berte  
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.: Ehlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.  
Koeber, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kuita, K.; Lapidis, A.; Ladiouis,  
A.: Authors: Lauber, J.; Lazarevic, V.; Lee, S.W.; Levine, A.; Liu, H.; Masuda, S.; Meue  
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033

A:Accession: H69828

A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-673 <RUN>  
A:Cross-references: GB:299109; GB:AL009126; NID:92633260; PIDN:CAB12811.1; PID:el182973;

A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yheH

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
C:Keywords: ATP; nucleotide binding; P-loop

C:Keywords: ATP; ATP-binding cassette homology <ABC>  
F:446-640/Domain: ATP-binding cassette homology

F:463-470/Region: nucleotide-binding motif A (P-loop)

Query Match  
Best Local Similarity 81.6%; Score 31; DB 2; Length 673;  
71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7  
||:||||  
D5 86 KEAHIYQ 92

## RESULT 12

B95172

hypothetical protein SPI477 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: B95172

R:Letellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: B95172

A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-135 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75571.1; PID:914972967; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPI477

Query Match  
Best Local Similarity 78.9%; Score 30; DB 2; Length 135;  
71.4%; Pred. No. 37;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7  
||:||||  
D5 89 KELYLYQ 95

## RESULT 13

B98038

hypothetical protein spr1331 [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: B98038

R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.  
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAlren, S.  
J.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:1154234

A:Accession: B98038

A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-135 <KUR>  
A:Cross-references: GB:AE007317; PIDN:JAL00135.1; PID:915458976; GSPDB:GN00174

C:Genetics:

A:Gene: spr1331

Query Match  
Best Local Similarity 78.9%; Score 30; DB 2; Length 135;  
71.4%; Pred. No. 37;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7  
||:||||  
D5 89 KELYLYQ 95

## RESULT 14

S74963

hypothetical protein slr1599 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S74963

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O., K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74963

A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-188 <KAN>  
A:Cross-references: EMBL:DB0902; GB:AB001339; NID:91652027; PIDN:BAI17003.1; PID:d101

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:



A:Start codon: GTG

## Query Match

78.9%; Score 30; DB 2; Length 188;

Best Local Similarity 83.3%; Pred. No. 52;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 EIHLYQ 7

|||||

DB 64 EIHLYE 69

## RESULT 15

E83420

endopeptidase Clp (EC 3.4.21.92) chain P PA1801 [similarity] - Pseudomonas aeruginosa (S

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-sep-2000 #text\_change 03-Aug-2001

C:Accession: E83420

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lofy, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; M01D:20437337

A:Accession: E83420

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-213 &lt;STO&gt;

A:Cross-references: GB:AE004606; GB:AE004091; NID:99947780; PIDN:AA05190.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: clpP; PA1801

C:Superfamily: endopeptidase Clp chain P

C:Keywords: hydrolase; serine proteinase

F:114/Active site: Ser #status predicted

F:139/Active site: His #status predicted

## Query Match

78.9%; Score 30; DB 2; Length 213;

Best Local Similarity 83.3%; Pred. No. 59;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLY 6

|||||

DB 74 KDTHLY 79

Search completed: August 20, 2002, 11:10:36  
Job time: 5666 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:00 ; Search time 55.29 seconds

(Without alignments)  
4.902 Million cell updates/sec

Title: US-09-824-286-14  
Perfect score: 38  
Sequence: 1 KEIHLXQ 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	38	100.0	369	1	CYRG_HUMAN
2	32	84.2	212	1	COAT_CYMV
3	31	81.6	51	1	YORO_TTV1
4	31	81.6	212	1	PSB2_ORYSA
5	31	81.6	294	1	PNK_VIBCH
6	31	81.6	373	1	CYRG_CANFA
7	31	81.6	472	1	MU49_YEAST
8	31	81.6	756	1	CPSB_DROME
9	30	78.9	50	1	REPA_BPT4
10	30	78.9	213	1	CLP1_PSEAE
11	30	78.9	215	1	COAT_PMY
12	30	78.9	357	1	HPPD_PSESP
13	30	78.9	4196	1	DYHC_SCHPO
14	29	76.3	223	1	YGB5_PASMU
15	29	76.3	286	1	YGB8_PASMU
16	29	76.3	380	1	CAGD_EDMIC
17	29	76.3	411	1	RAPS_CHICK
18	29	76.3	882	1	Y890_MYCTU
19	29	76.3	3674	1	SPCR_HUMAN
20	28	73.7	152	1	NDKH_MOUSE
21	28	73.7	197	1	COAE_BACNO
22	28	73.7	260	1	FAPR_ECOLI
23	28	73.7	359	1	PST_CRIGR
24	28	73.7	359	1	PST_HUMAN
25	28	73.7	359	1	PST_MOUSE
26	28	73.7	377	1	PRGR_SHEEP
27	28	73.7	415	1	CGA2_XENLA
28	28	73.7	431	1	T232_BACTB
29	28	73.7	433	1	APL5_HUMAN
30	28	73.7	581	1	SYP_CHLTR
31	28	73.7	628	1	V70K_TYMC
32	28	73.7	628	1	V70K_TYMC
33	28	73.7	656	1	YC26_PORPU

34	28	73.7	692	1	FHLA_ECOLI	P19323	escherichia
35	28	73.7	832	1	YFC4_YEAST	P43572	saccharomyc
36	28	73.7	852	1	YKML_YEAST	P2330	saccharomyc
37	28	73.7	923	1	PRGR_MOUSE	000175	mus musculu
38	28	73.7	923	1	PRGR_RAT	063449	rattus norv
39	28	73.7	930	1	PRGR_RABIT	P06186	oryctolagus
40	28	73.7	933	1	PRGR_HUMAN	P06401	homo sapien
41	28	73.7	1024	1	CARC_HUMAN	Q99P4	homo sapien
42	28	73.7	1045	1	SPS_BETVU	P49031	beta vulgar
43	28	73.7	2294	1	YCF2_ARATH	P56786	arabidopsis
44	27	71.1	90	1	HMCC_METSO	P15251	methanobact
45	27	71.1	101	1	TBCA_ARATH	004350	arabidopsis

## ALIGNMENTS

RESULT	1	STANDARD	PRT	369 AA.
CYRG_HUMAN				
ID	P31785			
AC	P31785			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cytokine receptor common gamma chain precursor (gamma-C) (interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64) (CD132 antigen).			
GN	IL2RG			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92355883; PubMed=1631559.			
RA	Takeshita T., Asao H., Ohtani K., Ishii N., Kumaki S., Tanaka N.,			
RA	Munakata H., Nakamura M., Sugamura K.;			
RT	"Cloning of the gamma chain of the human IL-2 receptor.";			
RL	Science 257:379-382(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=93293887; PubMed=8514792;			
RA	Noguchi M., Adelstein S., Cao X., Leonard W.J.;			
RT	"Characterization of the human interleukin-2 receptor gamma chain			
RL	gene.";			
RL	J. Biol. Chem. 268:13601-13608(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS ASP-114 AND ASN-153.			
RX	MEDLINE=94004847; PubMed=8401490;			
RA	Puck J.M., Deschenes S.M., Porter J.C., Dutra A.S., Brown C.J.,			
RA	Willard H., Henthorn P.S.;			
RT	"The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated			
RL	in X-linked severe combined immunodeficiency, SCIDX1.";			
RL	Hum. Mol. Genet. 2:1099-1104(1993).			
RN	[4]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94000315; PubMed=8266076;			
RA	Kondo M., Takeshita T., Ishii N., Nakamura M., Watanabe S.,			
RT	"Sharing of the interleukin-2 (IL-2) receptor gamma chain between			
RL	receptors for IL-2 and IL-4.";			
RL	Science 262:1874-1877(1993).			
RN	[5]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94090317; PubMed=8266076;			
RA	Russell S.M., Kegan A.D., Harada N., Nakamura Y., Noguchi M.,			
RA	Leonard P., Friedmann M.C., Miyajima A., Puri R.K., Paul W.E.,			
RA	Leonard W.J.;			
RT	"interleukin-2 receptor gamma chain: a functional component of the			
RL	interleukin-4 receptor.";			
RL	Science 262:1880-1883(1993).			
RN	[6]			
RP	IDENTIFICATION AS A IL-7R SUBUNIT.			

RX MEDLINE-94090316; PubMed-8266077;  
 RA Noguchi M., Nakamura Y., Russell S.M., Ziegler S.F., Tsang M., Cao X.,  
 RA Leonardi W.J.;  
 RT "Interleukin-2 receptor gamma chain: a functional component of the  
 RT Interleukin-7 receptor.";   
 RL Science 262:1877-1880(1993).  
 RN [7]  
 RP 3D-STRUCTURE MODELING OF 57-248.  
 RX MEDLINE-95111955; PubMed-7529123;  
 RA Bamorough P., Hedgecock C.J., Richards W.G.;  
 RT "The interleukin-2 and interleukin-4 receptors studied by molecular  
 RT modelling.";   
 RL Structure 2:839-851(1994).  
 RN [8]  
 RP VARIANTS XSCID PHE-115; CYS-240 AND ILE-241.  
 RX MEDLINE-94130970; PubMed-829698;  
 RA Disanto J.P., Dautry-Varsat A., Certain S., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Interleukin-2 (IL-2) receptor gamma chain mutations in X-linked  
 RT severe combined immunodeficiency disease result in the loss of  
 RT high-affinity IL-2 receptor binding.";   
 RL Eur. J. Immunol. 24:475-479(1994).  
 RN [9]  
 RP VARIANT XSCID IYS-68.  
 RX MEDLINE-94375038; PubMed-8088810;  
 RA Markiewicz S., Subtil A., Dautry-Varsat A., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Detection of three nonsense mutations and one missense mutation in  
 RT the Interleukin-2 receptor gamma chain gene in SCIDX1 that  
 RT differently affect the mRNA processing.";   
 RL Genomics 21:291-293(1994).  
 RN [10]  
 RP VARIANT XSCID HIS-162.  
 RX MEDLINE-94300093; PubMed-8027558;  
 RA Ishii N., Asao H., Kimura Y., Takeshita T., Nakamura M., Tsuchiya S.,  
 RA Kono T., Meeda M., Uchida Y., Sugamura K.;  
 RT "Impairment of ligand binding and growth signaling of mutant IL-2  
 RT receptor gamma-chains in patients with X-linked severe combined  
 RT immunodeficiency.";   
 RL J. Immunol. 153:1310-1317(1994).  
 RN [11]  
 RP VARIANT XSCID ASN-39.  
 RX MEDLINE-95023932; PubMed-7937790;  
 RA Disanto J.P., Rieux-Laucat F., Dautry-Varsat A., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Defective human Interleukin 2 receptor gamma chain in an atypical X  
 RT chromosome-linked severe combined immunodeficiency with peripheral T  
 RT cells.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9466-9470(1994).  
 RN [12]  
 RP VARIANTS XSCID CYS-226 AND HIS-226.  
 RX MEDLINE-95397841; PubMed-768284;  
 RA Pepper A.E., Buckley R.H., Small T.N., Puck J.M.;  
 RT "Two mutational hotspots in the interleukin-2 receptor gamma chain  
 RT gene causing human X-linked severe combined immunodeficiency.";   
 RL Am. J. Hum. Genet. 57:564-571(1995).  
 RN [13]  
 RP VARIANT XSCID SER-183.  
 RX MEDLINE-96013903; PubMed-7557965;  
 RA Clark P.A., Lester T., Genet S., Jones A.M., Hendriks R.,  
 RA Levinsky R.L., Kinon C.;  
 RT "Screening for mutations causing X-linked severe combined  
 RT immunodeficiency in the IL-2R gamma chain gene by single-strand  
 RT conformation polymorphism analysis.";   
 RL Hum. Genet. 96:427-432(1995).  
 RN [14]  
 RP VARIANT XSCID GLN-237 G-H-M INS.  
 RX MEDLINE-95164726; PubMed-7860773;  
 RA Puck J.M., Pepper A.E., Bedard P.-M., Laframboise R.;  
 RT "Female germ line mosaicism as the origin of a unique IL-2 receptor  
 RT gamma-chain mutation causing X-linked severe combined  
 RT immunodeficiency.";   
 RL J. Clin. Invest. 95:895-899(1995).

RN [15]  
 RP VARIANT XSCID GLN-293.  
 RX MEDLINE-95190013; PubMed-7883965;  
 RA Schmalstieg F.C., Leonard W.J., Noguchi M., Berg M., Rudloff H.E.,  
 RA Denney R.M., Dave S.K., Brooks E.G., Goldman A.S.;  
 RT "Missense mutation in exon 7 of the common gamma chain gene causes a  
 RT moderate form of X-linked combined immunodeficiency.";   
 RL J. Clin. Invest. 95:1169-1173(1995).  
 RN [16]  
 RP VARIANT XSCID ARG-115.  
 RX MEDLINE-97042245; PubMed-8900089;  
 RA Stephan V., Mahn V., Le Delst F., Dirksen U., Broeker B.,  
 RA Mueller-Fleckenstein I., Horneff G., Schroten H., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Atypical X-linked severe combined immunodeficiency due to possible  
 RT spontaneous reversion of the genetic defect in T cells.";   
 RL New Engl. J. Med. 335:1563-1567(1996).  
 RN [17]  
 RP VARIANT XSCID GLN-285.  
 RX MEDLINE-97295086; PubMed-9150740;  
 RA Jones A.M., Clark P.A., Katz F., Genet S., McMahon C., Alterman L.,  
 RA Cant A., Kinon C.;  
 RT "B-cell-negative severe combined immunodeficiency associated with a  
 RT common gamma chain mutation.";   
 RL Hum. Genet. 99:677-680(1997).  
 RN [18]  
 RP VARIANT XSCID CYS-222.  
 RX MEDLINE-98064061; PubMed-9399950;  
 RA Sharfe N., Shahar M., Rolfman C.M.;  
 RT "An interleukin-2 receptor gamma chain mutation with normal thymus  
 RT morphology.";   
 RL J. Clin. Invest. 100:3036-3043(1997).  
 CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF  
 CC INTERLEUKINS.  
 CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND  
 CC PROBABLY ALSO THE IL-13 RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A SEVERE COMBINED  
 CC IMMUNODEFICIENCY, WHICH IS KNOWN AS AGAMMAGLOBULINEMIA, SWISS TYPE  
 CC OR X-LINKED SEVERE COMBINED IMMUNODEFICIENCY DISEASE (XSCID).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- DATABASE: NAME=PROW: NOTE=CD guide CD132 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd132.htm".  
 CC -1- DATABASE: NAME=IL2Gbase; NOTE=X-linked SCID mutation database;  
 CC WWW="http://www.nhgrl.nih.gov/DIR/GMNB/SCID/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D11086; BAA01857.1; -;  
 DR EMBL; L12183; AAA59145.1; -;  
 DR EMBL; L12178; AAA59145.1; JOINED.  
 DR EMBL; L12176; AAA59145.1; JOINED.  
 DR EMBL; L12177; AAA59145.1; JOINED.  
 DR EMBL; L12179; AAA59145.1; JOINED.  
 DR EMBL; L12180; AAA59145.1; JOINED.  
 DR EMBL; L12181; AAA59145.1; JOINED.  
 DR EMBL; L12182; AAA59145.1; JOINED.  
 DR EMBL; L19546; AAC37524.1; -;  
 DR PIR; A42565; A42565;  
 DR PDB; 1TLN; 26-JAN-95.  
 DR MIM; 308380; -;  
 DR MIM; 300400; -;  
 DR InterPro; IPR002996; CR1A.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003531; Hematopo\_receptor\_S\_F1.

Query Match 100.0%; Score 38; DB 1; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 0.57; Mismatches 0; Indels 0; Gaps 0;  
 Matches 7; Conservative 0;

OY 1 KEIHLXQ 7  
 |||||  
 Db 120 KEIHLXQ 126

## RESULT 2

COAT\_CYWV STANDARD; PRT; 212 AA.  
 AC P16486;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DE 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Coat protein (Capsid protein).  
 OS Clover yellow mosaic virus (CYMV).  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.  
 CC NCBI\_TaxID=12177;

RP SEQUENCE FROM N.A.  
 RX MEDLINE=89293092; PubMed=2738582;  
 RA Abouhaidar M.G., Lai R.;  
 RT "Nucleotide sequence of the 3'-terminal region of clover yellow  
 RT mosaic virus RNA";  
 RL J. Gen. Virol. 70:1871-1875(1989).

CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS  
 CC PARTICLES.  
 CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D00485; BAA00375.1; -  
 DR InterPro: IPR000052; Potex\_carlavirus-coat.  
 DR Pfam: PF00286; virus\_p-coat; 1.  
 DR PRINTS: PR00232; POTXCARLCOAT.  
 DR ProDom: PD00603; Potex\_carlavirus-coat; 1.  
 DR PROSITE: PS00418; POTEX\_CARLAVIRUS\_COAT; 1.  
 KW Coat protein.  
 SQ SEQUENCE 212 AA; 23438 MW; 30D6B8773F5D6E17 CRC64;

Query Match 84.2%; Score 32; DB 1; Length 212;  
 Best Local Similarity 71.4%; Pred. No. 6.4; Mismatches 1; Indels 0; Gaps 0;  
 Matches 5; Conservative 1;

OY 1 KEIHLXQ 7  
 |||||  
 Db 171 KNVHLXQ 177

## RESULT 3

YORO\_TTV1 STANDARD; PRT; 51 AA.  
 AC P19299;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE Hypothetical 6.2 kDa protein.  
 OS Thermoplasma tenax virus 1 (strain KRA1) (TTV1).  
 CC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;  
 CC Lipothirixvirus.  
 CC NCBI\_TaxID=10480;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Neumann H.;  
 RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X1485; CAA32995.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 51 AA; 6171 MW; 7B2960FEFF195754 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 51;  
 Best Local Similarity 83.3%; Pred. No. 2.3; Mismatches 1; Indels 0; Gaps 0;  
 Matches 5; Conservative 1;

OY 2 EIHLYQ 7  
 |||||  
 Db 5 EIHLYQ 10

## RESULT 4

PSB2\_ORYSA STANDARD; PRT; 212 AA.  
 ID PSB2\_ORYSA  
 AC Q9LST6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proteasome subunit beta type 2 (EC 3.4.25.1) (20S proteasome alpha  
 DE subunit D) (20S proteasome subunit beta-4).  
 GN PBD1.  
 OS Oryza sativa (Rice).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriophytidae; Oryzaceae; Oryza.  
 OC NCBI\_TaxID=4530;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RX MEDLINE=20314477; PubMed=10854779;  
 RA Sassa H., Oguchi S., Inoue T., Hirano H.;  
 RT "Primary structural features of the 20S proteasome subunits of rice  
 RT (Oryza sativa).";  
 RL Gene 250:61-66(2000).

CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX  
 CC WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,  
 CC PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR  
 CC SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC  
 CC ACTIVITY (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad  
 CC specificity.

CC -1- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal  
 CC proteolytic pathway.

CC -1- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL  
 CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE (BY  
 CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B; ALSO KNOWN AS THE  
 CC PROTEASOME B-TYPE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB026566; BAA96837.1; -

DR InterPro: IPR001353; Proteasome.  
 DR InterPro: IPR000243; Proteasome\_B.  
 DR Pfam: PF00227; Proteasome; 1.  
 DR PROSITE: PS00854; PROTEASOME\_B; FALSE\_NEG.  
 KM Proteasome: Hydrolyase; Protease.  
 SO SEQUENCE 212 AA; 23444 MW; 9A87F8050078FF9 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 212;  
 Best Local Similarity 71.4%; Pred. No. 11;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KEHLQ 7  
 | : ||| |  
 Db 62 KNLHLQ 68

RESULT 5  
 ID PPNK\_VIBCH STANDARD; PRT; 294 AA.  
 AC Q9KTP8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)  
 DE (Poly(P)/ATP NAD kinase).  
 GN PPNK OR VC0853.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Givann M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Uitterback T., Fleischmann J.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "NNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae";  
 RL Nature 406:477-483(2000).  
 CC -1- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes  
 ATP and other nucleoside triphosphates as well as inorganic  
 CC polyphosphate as a source of phosphorus (By similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+).  
 CC -1- COFACTOR: Requires divalent metal ions for activity (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004170; AAF94015.1; -.  
 DR TIGR: VC0853; -.  
 DR InterPro: IPR002504; DUF15.  
 DR Pfam: PF01513; DUF15; 1.  
 KM Transferase: Kinase; NAD: Complete proteome.  
 SO SEQUENCE 294 AA; 32698 MW; 7FEAA38920AAAE8 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 294;  
 Best Local Similarity 83.3%; Pred. No. 15;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 EHLQ 7  
 ||| : ||  
 Db 259 EHLQ 264

RESULT 6  
 ID CYRG\_CANFA STANDARD; PRT; 373 AA.  
 AC P40321;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Cytokine receptor common gamma chain precursor (gamma-C)  
 DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (p64).  
 GN IL2RG.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RX MEDLINE=95110114; PubMed=7829104;  
 RA Henthorn P.S., Somberg R.L., Fimiani V.M., Puck J.M., Patterson D.F.,  
 RA Felsburg P.J.;  
 RT "IL-2R gamma gene microdeletion demonstrates that canine X-linked  
 RT severe combined immunodeficiency is a homologue of the human  
 RT disease";  
 RL Genomics 23:69-74(1994).  
 CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF  
 CC INTERLEUKINS.  
 CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND  
 CC PROBABLY ALSO THE IL-13 RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A CANINE X-LINKED  
 CC SEVERE COMBINED IMMUNODEFICIENCY.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U04361; AAC48403.1; -.  
 DR HSP, P31785; IILN.  
 DR InterPro: IPR002996; CRIA.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003531; Hematopo\_receptor\_S\_Fl.  
 DR Pfam: PF00041; fn3; 1.  
 DR SMART: SM00060; FN3; 1.  
 DR PROSITE: PS01355; HEMATOPO\_REC\_S\_Fl; 1.  
 KM Receptor: Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 373  
 FT DOMAIN 23 261  
 FT TRANSSEM 262 283  
 FT DOMAIN 284 373  
 FT DOMAIN 151 249  
 FT DOMAIN 151 249  
 FT DISULFID 62 72  
 FT DISULFID 102 115  
 FT CARBOHYD 24 24  
 FT CARBOHYD 71 71  
 FT CARBOHYD 75 75  
 FT CARBOHYD 84 84  
 FT CARBOHYD 159 159  
 FT CARBOHYD 164 164  
 FT CARBOHYD 249 249  
 SO SEQUENCE 373 AA; 42516 MW; 03A0DE1F8B089DBB CRC64;

Query Match 81.6%; Score 31; DB 1; Length 373;  
 Best Local Similarity 71.4%; Pred. No. 19;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLVQ 7  
 :|||||:  
 Db 120 EEIHLVE 126

## RESULT 7

NU49\_YEAST STANDARD; PRT; 472 AA.  
 ID NU49\_YEAST 002199;  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nucleoporin NUP49/NSP49 (Nuclear pore protein NUP49/NSP49).  
 GN NUP49 OR NSP49 OR YGL172W OR G1648.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9309880; PubMed=1464327;  
 RA Wimmer C., Doye V., Grandi P., Nehrbass U., Hurt E.C.:  
 RT "A new subclass of nucleoporins that functionally interact with  
 RT nuclear pore protein NSP1.".  
 RL EMBO J. 11:5051-5061(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93054906; PubMed=1385442;  
 RA Menten S.R., Rout M.P., Blobel G.:  
 RT "A new family of yeast nuclear pore complex proteins.".  
 RL J. Cell Biol. 119:703-723(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / FY1679;  
 RX MEDLINE=96109931; PubMed=8619317;  
 RA Bertani L., Coglievina M., Zaccaria P., Klita R., Bruschi C.V.:  
 RT "The sequence of an 11.1 kb fragment on the left arm of Saccharomyces  
 RT cerevisiae chromosome VII reveals six open reading frames including  
 RT NSP49, KEM1 and four putative new genes.".  
 RL Yeast 11:1187-1194(1995).  
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.  
 CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF  
 CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.  
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.  
 CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X68109; CAA48229.1; -  
 DR EMBL: 215040; CAA78758.1; -  
 DR EMBL: X84705; CAA59181.1; -  
 DR EMBL: 272694; CAA96884.1; -  
 DR PIR: S28026; S28026;  
 DR PIR: S28538; S28538;  
 DR PIR: C44402; C44402;  
 DR SGD: S0003140; NUP49.  
 KW Nuclear protein; Transport; Coiled coil; Repeat.  
 FT DOMAIN 14 236 14 X 6 AA APPROXIMATE REPEATS OF  
 FT G-L-F-G.  
 FT DOMAIN 350 472 COILED COIL (POTENTIAL).  
 FT GIBBS R.A., Myers E.W., Rubin G.M., Venter J.C.:  
 SQ SEQUENCE 472 AA; 49142 MW; 0CAV516FF5753CA3 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 472;  
 Best Local Similarity 71.4%; Pred. No. 25;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLVQ 7  
 :|||||:  
 Db 375 KTHLVE 381

## RESULT 8

CPSB\_DROME STANDARD; PRT; 756 AA.  
 ID CPSB\_DROME 09V3D6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Probable cleavage and polyadenylation specificity factor, 100 kDa  
 DE subunit (CPSP 100 kDa subunit).  
 GN BCDNA:ID14168 OR CG1957.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,  
 RA Lewis S.E., Suh C., Rubin G.M.:  
 RT "Full length Drosophila melanogaster cDNA sequence.".  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBD databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKLEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,  
 RA Wan R.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Planck C., Baldwin D.,  
 RA Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostel A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostrelli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
 RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: CPSE PLAYS A KEY ROLE IN PRE-MRNA 3'-END FORMATION,  
 CC RECOGNIZING THE AUAUAA SIGNAL SEQUENCE AND INTERACTING WITH  
 CC POLY(A) POLYMERASE AND OTHER FACTORS TO BRING ABOUT CLEAVAGE AND  
 CC POLY(A) ADDITION (BY similarity).  
 CC -1- SUBUNIT: CPSE IS A HETEROTETRAMER COMPOSED OF FOUR DISTINCT  
 CC SUBUNITS 160, 100, 70 AND 30 kDa (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE CPSE100 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF160933; AAD46873.1; -.  
 DR EMBL: AE003768; AAF56844.1; -.  
 DR FLYBase: FBgn0027873; BCDNA:LD14168.  
 KW mRNA processing; Nuclear protein; RNA-binding.  
 SQ SEQUENCE 756 AA; 85418 MW; E391D61CFADDD821 CRC64;

Query Match 81.6%; Score 31; DB 1; Length 756;  
 Best Local Similarity 83.3%; Pred. No. 41;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2 EIHLYQ 7  
 11111  
 Db 611 EIHLYQ 616

RESULT 9  
 REPA\_BPT4 STANDARD; PRT; 50 AA.  
 ID P32284;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE DNA replication protein repA (DNA-binding protein dbpB).  
 GN REPEA OR DBP.  
 OS Bacteriophage T4.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like phages.  
 OX NCBI\_TaxID=10665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91190815; PubMed-2488704.  
 RA Mosig G., Lin G.W., Franklin J., Fan W.H.;  
 RT "Functional relationships and structural determinants of two  
 RT bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-  
 RT associated (gene 5) protein.";  
 RL New Biol. 1:171-179(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,  
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;  
 RT "Bacteriophage T4 genome analysis.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP IDENTIFICATION, AND FUNCTION.  
 RA MEDLINE-20026826; PubMed-10559179;  
 RA Valiskunne R., Miller A., Davenport L., Mosig G.;  
 RT "Two new early bacteriophage T4 genes, repA and repB, that are  
 RT important for DNA replication initiated from origin E.";  
 RL J. Bacteriol. 181:7115-7125(1999).  
 CC -1- FUNCTION: Involved in T4 DNA replication. Binds to ssDNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X15728; -. NOT\_ANNOTATED\_CDS.  
 DR EMBL: AF158101; AAD42504.1; -.  
 KW DNA-binding; DNA replication.  
 SQ SEQUENCE 50 AA; 6130 MW; 33AA9EB42026EF6 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 50;  
 Best Local Similarity 83.3%; Pred. No. 3.7;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 KEIHLX 6  
 11111  
 Db 39 KDIHLX 44

RESULT 10  
 CLP1\_PSEAE STANDARD; PRT; 213 AA.  
 ID Q912U1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ATP-dependent Clp protease proteolytic subunit 1 (EC 3.4.21.92)  
 DE (Endopeptidase Clp 1).  
 GN CLP1 OR PA1801.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 15692 / PA01;  
 RX MEDLINE-20437337; PubMed-10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Medman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larrbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: Cleaves peptides in various proteins in a process that  
 CC requires ATP hydrolysis. Has a chymotrypsin-like activity. Plays a  
 CC major role in the degradation of misfolded proteins (By  
 CC similarity).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins to small peptides in  
 CC the presence of ATP and magnesium. Alpha-casein is the usual test  
 CC substrate. In the absence of ATP, only oligopeptides shorter than  
 CC five residues are cleaved (such as succinyl-leu-tyr-l-NMEC; and  
 CC Leu-tyr-leu-l-tyr-tyr, in which the cleavage of the -Tyr-l-Leu-  
 CC and -Tyr-l-Tyr-bond also occurs).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S14; ALSO KNOWN AS CLPP  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE004606; AAC05190.1; -.  
 DR InterPro: IPR001907; Clp\_protease.  
 DR Pfam: PF00574; Clp\_protease; 1.  
 DR PRINTS: PR00127; CLPPROTEASEP.



DR PROSITE; PS00382; CLP\_PROTEASE\_HIS; 1.  
 DR PROSITE; PS00381; CLP\_PROTEASE\_SER; 1.  
 KM Hydrolase; Serine protease; Complete proteome.  
 FT ACT\_SITE 114 114 BY SIMILARITY.  
 FT ACT\_SITE 139 139 BY SIMILARITY.  
 SQ SEQUENCE 213 AA; 23502 MW; ED5A7365E5A85E3 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 213;  
 Best Local Similarity 83.3%; Pred. No. 18;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 6  
 Db 74 KDHLX 79

## RESULT 11

COAT\_PMV STANDARD; PRT; 215 AA.  
 ID COAT\_PMV  
 AC P16596;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Coat protein (Capsid protein).  
 OS Papaya mosaic potexvirus (PMV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.  
 OX NCBI\_TaxID=12181;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89381685; PubMed=2778435;  
 RA Slt T.L., Abouhalaid M.G., Holy S.;  
 RT "Nucleotide sequence of papaya mosaic virus RNA.";  
 RL J. Gen. Virol. 70:2325-2331(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88089532; PubMed=3335832;  
 RA Abouhalaid M.G.;  
 RT "Nucleotide sequence of the capsid protein gene and 3' non-coding  
 region of papaya mosaic virus RNA.";  
 RL J. Gen. Virol. 69:219-226(1988).  
 RN [3]  
 RP SEQUENCE.  
 RA Short M.N., Turner D.S., March J.F., Papin D.J.C., Parente A.,  
 RA Davies J.W.;  
 RT "The primary structure of papaya mosaic coat protein.";  
 RL Virology 152:280-283(1986).  
 CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS  
 PARTICLES.  
 CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 CC EMBL; D00240; BAA00169.1; -;  
 DR EMBL; D13957; BAA03054.1; -;  
 DR PIR; J00100; VCMGPM.  
 DR InterPro: IPR000052; Potex.carlavirius\_coat.  
 DR Pfam: PF00286; virus\_p-coat.1.  
 DR PRINTS; PR00232; POTXCARLCOAT.  
 DR ProDom: PD000603; Potex.carlavirius\_coat; 1.  
 DR PROSITE; PS00418; POTEX\_CARLAVIRIUS\_COAT; 1.  
 KM Coat protein.  
 FT .CONFLICT 1 5 MISSING (IN REF. 3).  
 FT .CONFLICT 11 11 T -> I (IN REF. 1).  
 FT .CONFLICT 30 30 S -> SN (IN REF. 3).  
 FT .CONFLICT 77 77 MISSING (IN REF. 2).  
 FT .CONFLICT 185 185 Q -> E (IN REF. 3).  
 RN CONFLICT

FT CONFLICT 190 190 T -> A (IN REF. 3).  
 FT CONFLICT 215 215 E -> Q (IN REF. 3).  
 SQ SEQUENCE 215 AA; 23033 MW; 605B94EE70B41323 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 215;  
 Best Local Similarity 57.1%; Pred. No. 18;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 7  
 Db 175 KQVHLFQ 181

## RESULT 12

HPPD\_PSESP STANDARD; PRT; 357 AA.  
 ID HPPD\_PSESP  
 AC P80064;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27) (4HPPD) (HPD).  
 OS Pseudomonas sp. (strain P.J. 874).  
 OC Bacteria; Proteobacteria.  
 OX NCBI\_TaxID=306;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=92241278; PubMed=1572351;  
 RA Ruetschli U., Odelhoeg B., Lindstedt S., Barros-Soederling J.,  
 RA Petersen B., Joernvall H.;  
 RT "Characterization of 4-hydroxyphenylpyruvate dioxygenase. Primary  
 structure of the Pseudomonas enzyme.";  
 RL Eur. J. Biochem. 205:459-466(1992).  
 CC -1- CATALYTIC ACTIVITY: 4-hydroxyphenylpyruvate + O(2) = homogentisate  
 + CO(2).  
 CC -1- COFACTOR: IRON.  
 CC -1- PATHWAY: CATABOLISM OF TYROSINE; SECOND STEP, CATABOLISM OF  
 PHENYLALANINE; THIRD STEP.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SIMILARITY: BELONGS TO THE 4HPPD FAMILY.  
 DR PIR; S21209; S21209.  
 DR InterPro: IPR000325; Glyoxalase\_1.  
 DR Pfam: PF00903; Glyoxalase; 1.  
 KM Oxidoreductase; Dioxygenase; Iron; Phenylalanine catabolism;  
 KM Tyrosine catabolism.  
 FT DOMAIN 167 196 TYR-RICH.  
 SQ SEQUENCE 357 AA; 40060 MW; 26CF480B1484BD0 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 357;  
 Best Local Similarity 57.1%; Pred. No. 31;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLX 7  
 Db 44 KDVLXR 50

RESULT 13  
 DYHC\_SCHPO STANDARD; PRT; 4196 AA.  
 ID DYHC\_SCHPO  
 AC O13290; O9P6L0; Q9UTP8;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Dynein heavy chain, cytosolic (DYHC).  
 GN DHCI OR SPAC30C2.01C OR SPAC1093.06C.  
 OS Schizosaccharomyces pombe (Fission Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-CRL152;
RA Yamamoto A., West R.R., McIntosh J.R., Hiraoka Y.;
RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE OF 1-2340 FROM N.A.
RC STRAIN-972;
RA Saunders D., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RN Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 2308-4196 FROM N.A.
RC STRAIN-972;
RA Seeger K., Harris D., McDougall R.C., Rajandream M.A., Barrell B.G.;
RN Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RL [1]
CC FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES. REQUIRED FOR NUCLEAR MOVEMENT
CC DURING MEIOTIC PROPHASE.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROBABLY BINDS INDIRECTLY TO
CC THE INNER PLASMA MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB006784; BAA22056.1; -
DR EMBL: AL353652; CAB90788.1; -
DR EMBL: AL132839; CAB60251.1; -
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 1217 1252 COILED COIL (POTENTIAL).
FT DOMAIN 1984 2012 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 3315 3403 COILED COIL (POTENTIAL).
FT DOMAIN 3649 3666 COILED COIL (POTENTIAL).
FT NP_BIND 1890 1897 ATP (POTENTIAL).
FT NP_BIND 2169 2176 ATP (POTENTIAL).
FT NP_BIND 2174 2181 ATP (POTENTIAL).
FT NP_BIND 2520 2527 ATP (POTENTIAL).
SQ SEQUENCE 4196 AA; 484308 MW; 8F10AE370184FC0C CRC64;

Query Match 78.9%; Score 30; DB 1; Length 4196;
Best Local Similarity 71.4%; Pred. No. 4,3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLXQ 7
Db 1320 EEIHLXK 1326

RESULT 14
YGBS_PASMU STANDARD; PRT; 233 AA.
AC O9CKD7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein PM1685.
GN PM1685.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PM70;
RX MEDLINE=21145866; PubMed=11248100;

```

```

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RL "Complete genomic sequence of Pasteurella multocida pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- SIMILARITY: BELONGS TO THE PIRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006205; AAK03769.1; -
DR InterPro: IPR003829; DUF209.
DR Pfam: PF02678; DUF209; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26804 MW; 28500A6C859F65D5 CRC64;

Query Match 76.3%; Score 29; DB 1; Length 233;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 EIHLYQ 7
Db 113 DVHLYQ 118

RESULT 15
YGBB_EDWIC STANDARD; PRT; 286 AA.
ID YGBB_EDWIC
AC O52401;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 30.6 kDa protein in IC1A-FBA intergenic region.
GN YGBB.
OS Edwardsiella ictaluri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Edwardsiella.
OX NCBI_TaxID=67780;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-93-146;
RA Moore M.M., Fernandez D.H., Thune R.L.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0003 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF037440; AAB92571.1; -
DR InterPro: IPR001880; UPF0003.
DR Pfam: PF00924; MS channel; 1.
DR PROSITE: PS01246; UPF0003; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
SQ SEQUENCE 286 AA; 30648 MW; DB2442E07C70835D CRC64;

Query Match 76.3%; Score 29; DB 1; Length 286;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 EIHLYQ 7  
: : : : :  
Db 274 DVHLYQ 279

Search completed: August 20, 2002, 11:33:01  
Job time: 1450 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:53 ; Search time 191.14 Seconds  
(without alignments)  
6.335 Million cell updates/sec

Title: US-09-824-286-14  
Perfect score: 38  
Sequence: 1 KEIRLYQ 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	86.8	1611	10 Q9LKW5	Q9LKW5 aegilops ta
2	33	86.8	1628	10 Q9LKW6	Q9LKW6 triticum ae
3	32	84.2	257	12 Q6K343	Q6K343 clover yell
4	32	84.2	275	4 Q9H285	Q9H285 homo sapien
5	32	84.2	286	12 Q6K342	Q6K342 clover yell
6	32	84.2	312	12 Q6K339	Q6K339 clover yell
7	32	84.2	367	4 Q96EL4	Q96EL4 homo sapien
8	32	84.2	434	2 Q9JIS0	Q9JIS0 salmonella
9	32	84.2	551	10 Q9SVB6	Q9SVB6 arabidopsi
10	31	81.6	184	5 Q7J159	Q7J159 tetrahymena
11	31	81.6	233	16 P7J155	P7J155 synecocyst
12	31	81.6	506	10 Q9FHY9	Q9FHY9 arabidopsi
13	31	81.6	591	5 Q17412	Q17412 caenorhabdi
14	31	81.6	634	16 Q9CNH8	Q9CNH8 pasteurella
15	31	81.6	673	16 Q07549	Q07549 bacillus su
16	31	81.6	677	5 Q9VEL0	Q9VEL0 drosophila

17	31	81.6	785	5 Q9V089	Q9V089 drosophila
18	31	81.6	867	5 Q95RC2	Q95RC2 drosophila
19	30	78.9	135	16 Q97PW3	Q97PW3 streptococ
20	30	78.9	188	16 P72984	P72984 synecocyst
21	30	78.9	191	2 Q9AKB0	Q9AKB0 rickettsia
22	30	78.9	223	16 Q9CHAB	Q9CHAB lactococcus
23	30	78.9	244	2 Q34250	Q34250 wolfinella s
24	30	78.9	255	12 Q919K4	Q919K4 culix nigri
25	30	78.9	292	16 Q989S3	Q989S3 rhizobium l
26	30	78.9	319	17 Q97UJ8	Q97UJ8 sulfolobus
27	30	78.9	433	2 P94847	P94847 helicobacte
28	30	78.9	433	16 Q25321	Q25321 helicobacte
29	30	78.9	433	16 Q9ZLN2	Q9ZLN2 helicobacte
30	30	78.9	456	3 Q9US58	Q9US58 schizosacch
31	30	78.9	480	10 Q9M5C4	Q9M5C4 euphorbia e
32	30	78.9	482	5 Q9U0J4	Q9U0J4 plasmodium
33	30	78.9	553	5 Q9NFK4	Q9NFK4 boophilus m
34	30	78.9	553	5 Q9NFK3	Q9NFK3 boophilus m
35	30	78.9	553	5 Q9NFK2	Q9NFK2 boophilus m
36	30	78.9	553	5 Q9NFK1	Q9NFK1 boophilus m
37	30	78.9	752	12 Q913V5	Q913V5 baboon gamm
38	30	78.9	762	2 P96793	P96793 lactobacill
39	30	78.9	833	3 Q96V83	Q96V83 metarhizium
40	30	78.9	940	16 Q9KR83	Q9KR83 vibrio chol
41	30	78.9	1220	15 Q41894	Q41894 bovine sync
42	30	78.9	1322	10 Q9PLN3	Q9PLN3 arabidopsi
43	29	76.3	129	16 Q9PEU1	Q9PEU1 xyella fas
44	29	76.3	146	5 Q18058	Q18058 caenorhabdi
45	29	76.3	153	16 Q25338	Q25338 helicobacte

#### ALIGNMENTS

RESULT 1  
ID Q9LKW5 PRELIMINARY; PRT; 1611 AA.  
AC Q9LKW5;  
DT 01-OCT-2000 (TRENBLREL. 15, Created)  
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)  
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
DE STARCH SYNTHASE III.  
OS Aegilops tauschii (Aegilops squarrosa).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
OC Triticeae; Aegilops.  
OX NCBI\_TaxID=37682;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20317203; PubMed=10859191;  
RX Li Z., Mouille G., Kosar-Hashemi B., Rahman S., Clarke B., Gale K.R.,  
RA Appels R., Morell M.K.;  
RT "The structure and expression of the wheat starch synthase III gene.  
RT Motifs in the expressed gene define the lineage of the starch synthase  
RT III gene family.";  
RL Plant Physiol. 123:613-624(2000).  
DR EMBL: AF258609; AAF88000.1; -;  
SQ SEQUENCE 1611 AA; 180653 MW; 78BEC99C03FEAF9 CRC64;

Query Match 86.8%; Score 33; DB 10; Length 1611;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIRLYQ 7  
DB 1215 KDLRLYQ 1221  
RESULT 2  
ID Q9LKW6 PRELIMINARY; PRT; 1628 AA.  
AC Q9LKW6;

DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE STARCH SYNTHASE III.  
OS Triticum aestivum (Wheat).  
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Pooidae:  
OC Triticeae: Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
FX MEDLINE=20317203; PubMed=10859191;  
RA Li Z., Mouille G., Koser-Hashemi B., Rahman S., Clarke B., Gale K.R.,  
RA Appels R., Morell M.K.;  
RT "The structure and expression of the wheat starch synthase III gene.  
RT Motifs in the expressed gene define the lineage of the starch synthase  
RT III gene family.";  
RL Plant Physiol. 123:613-624(2000).  
DR EMBL: AF258608; AAF87999.1;  
SQ SEQUENCE 1628 AA; 183138 MW; 069498ADA859B819 CRC64;

Query Match 86.8%; Score 33; DB 10; Length 1628;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLQ 7  
Db 1232 KDLHLQ 1238

RESULT 3  
Q66343 PRELIMINARY; PRT; 257 AA.  
ID 066343;  
AC 066343;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE COAT PROTEIN (CAPSID PROTEIN).  
OS Clover yellow mosaic virus (CYMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.  
OX NCBI\_TaxID=12177;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8923092; PubMed=2738582;  
RA Abouhaidar M.G., Lai R.;  
RT "Nucleotide sequence of the 3'-terminal region of clover yellow mosaic  
RT virus RNA.";  
RL J. Gen. Virol. 70:1871-1875(1989).  
CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.  
DR EMBL: D00485; BAAD0373.1; -  
DR InterPro: IPR000052; Potex\_carlaviruses\_coat.  
DR InterPro: IPR000531; TONB\_BOX.  
DR Pfam: PF00286; virus\_P-coat; 1.  
DR PRINTS: PR00232; POTXCARCOAT.  
DR ProDom: PD000603; Potex\_carlavirus\_coat; 1.  
DR PROSITE: PS00418; POTEX\_CARLAVIRUS\_COAT; 1.  
DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; UNKNOWN\_1.  
KM Coat protein.  
SQ SEQUENCE 257 AA; 28095 MW; 4D9FA346DE9CACC6 CRC64;

Query Match 84.2%; Score 32; DB 12; Length 257;  
Best Local Similarity 71.4%; Pred. No. 47;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KEIHLQ 7  
Db 216 KNLHLQ 222

RESULT 4  
Q9H285 PRELIMINARY; PRT; 275 AA.  
ID 09H285;  
AC 09H285;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE SEROLOGICALLY DEFINED BREAST CANCER ANTIGEN NY-BR-24 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;  
OC Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BREAST;  
RA Scalan M.J., Gout I., Stockert E., Gure A.O., Jaeger D., Chen Y.-T.,  
RA Old L.J.;  
RT "Humoral Immunity to Human Breast Cancer: Antigen Definition and  
RT Quantitative Analysis of mRNA Expression.";  
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF308288; AAC48256.1; -  
DR InterPro: IPR000038; GTP\_Cell\_Div.  
DR Pfam: PF00735; GTP\_CDC; 1.  
DR ProDom: PD002565; GTP\_Cell\_Div; 1.  
FT NON\_TER 1  
SQ SEQUENCE 275 AA; 31893 MW; 517068962B98C7B1 CRC64;

Query Match 84.2%; Score 32; DB 4; Length 275;  
Best Local Similarity 71.4%; Pred. No. 50;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEIHLQ 7  
Db 102 EIHLYQ 108

RESULT 5  
Q66342 PRELIMINARY; PRT; 286 AA.  
ID 066342;  
AC 066342;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE COAT PROTEIN (CAPSID PROTEIN).  
OS Clover yellow mosaic virus (CYMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.  
OX NCBI\_TaxID=12177;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91306431; PubMed=1830181;  
RA White K.A., Bancroft J.B., Mackie G.A.;  
RT "Defective RNAs of clover yellow mosaic virus encode  
RT nonstructural/coat protein fusion products.";  
RL Virology 183:479-486(1991).  
CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.  
DR EMBL: M63514; AAA42938.1; -  
DR InterPro: IPR000052; Potex\_carlaviruses\_coat.  
DR Pfam: PF00286; virus\_P-coat; 1.  
DR ProDom: PD000603; Potex\_carlavirus\_coat; 1.  
KM Coat protein.  
SQ SEQUENCE 286 AA; 31730 MW; 9BE5B01BD1E32895 CRC64;

Query Match 84.2%; Score 32; DB 12; Length 286;  
Best Local Similarity 71.4%; Pred. No. 52;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KEIHLQ 7  
Db 245 KNLHLQ 251

RESULT 6  
 ID 066339 PRELIMINARY; PRT; 312 AA.  
 AC 066339;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE COAT PROTEIN (CAPSID PROTEIN)  
 OS Clover yellow mosaic virus (CYMV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage: Potexvirus.  
 NCBI\_TaxID=12177;  
 RX MEDLINE=91306431; PubMed=1830181;  
 RP SEQUENCE FROM N.A.  
 RA White K.A., Bancroft J.B., Mackie G.A.;  
 RT "Defective RNAs of clover yellow mosaic virus encode nonstructural/coat protein fusion products."  
 RL Virology 183:479-486(1991).  
 CC -1- FUNCTION: SELF-ASSEMBLES WITH THE RNA TO FORM INFECTIOUS PARTICLES (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE COAT PROTEINS OF OTHER POTEXVIRUSES.  
 DR EMBL; M63511; AAA2935.1; -.  
 DR Interpro: IPR000052; Potex\_carlavirius\_coat.  
 DR Pfam: PF00286; virus\_p-coat; 1.  
 DR ProDom: PD000603; Potex\_carlavirius\_coat; 1.  
 DR ProSITE: PS00418; POTEX\_CARLAVIRIUS\_COAT; 1.  
 KW Coat protein.  
 SQ SEQUENCE 312 AA; 34657 MW; BA17BDFB3E675E21 CRC64;

Query Match 84.2%; Score 32; DB 12; Length 312;  
 Best Local Similarity 71.4%; Pred. No. 57;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLYO 7  
 I :||||  
 DB 271 KNYHLYO 277

RESULT 7  
 ID 096EL4 PRELIMINARY; PRT; 367 AA.  
 AC 096EL4;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE SIMILAR TO SEPTIN 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PRIMARY B-CELLS FROM TONSILS;  
 RL Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC012161; AAH12161.1; -.  
 SQ SEQUENCE 367 AA; 41970 MW; E9CGBEF373CB0FC4 CRC64;

Query Match 84.2%; Score 32; DB 4; Length 367;  
 Best Local Similarity 71.4%; Pred. No. 67;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLYO 7  
 I :|||||  
 DB 194 ERIHIYO 200

RESULT 8  
 O93ISO

ID O93ISO PRELIMINARY; PRT; 434 AA.  
 AC O93ISO;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE SCIP PROTEIN.  
 GN SCIP.  
 OS Salmonella enterica subsp. enterica serovar Typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=90371;  
 RX MEDLINE=99348391; PubMed=10417651;  
 RP SEQUENCE FROM N.A.  
 RA Folkesson A.;  
 RC STRAIN-SR-11;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SR-11;  
 RX MEDLINE=99348391; PubMed=10417651;  
 RA Folkesson A., Advani A., Sukupolvi S., Pfeifer J.D., Normark S., Loeffel S.;  
 RT "Multiple insertions of fimbrial operons correlate with the evolution of Salmonella serovars responsible for human disease."  
 RL MOL. Microbiol. 33:612-622(1999).  
 DR EMBL; AJ320483; CAC48213.1; -.  
 SQ SEQUENCE 434 AA; 47355 MW; 3D787A09B8A092D CRC64;

Query Match 84.2%; Score 32; DB 2; Length 434;  
 Best Local Similarity 71.4%; Pred. No. 79;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLYO 7  
 I :|||||  
 DB 96 REMHLYO 102

RESULT 9  
 ID O9SVE6 PRELIMINARY; PRT; 551 AA.  
 AC O9SVE6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
 DE PUTATIVE GROWTH REGULATOR PROTEIN.  
 GN F22113.160 OR AT4G38390.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Medler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Medler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035539; CAB37495.1; -.  
 DR EMBL; AL161593; CAB80504.1; -.  
 DR Interpro: IPR004348; DUF246.

DR Pfam; PF03138; DUF246; 1.  
SQ SEQUENCE 551 AA; 62872 MW; 61FC1B84BA3756FC CRC64;  
Query Match  
Best Local Similarity 84.2%; Score 32; DB 10; Length 551;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KEIHLX 6  
||:||||  
Db 360 KEVHLX 365  
RESULT 10  
077159 PRELIMINARY; PRT; 184 AA.  
AC 077159;  
DT 01-NOV-1998 (TREMblrel\_08, Created)  
DT 01-NOV-1998 (TREMblrel\_08, Last sequence update)  
DT 01-DEC-2001 (TREMblrel\_19, Last annotation update)  
DE HETEROCHROMATIN-ASSOCIATED PROTEIN 1-LIKE PROTEIN.  
GN HHP1.  
OS Tetrahymena thermophila.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
OC Tetrahymenida; Tetrahymena.  
OX NCBI\_Taxid=5911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CU428;  
RX MEDLINE=99030619; PubMed=9811850;  
RA Huang H., Wiley E.A., Lending C.R., Allis C.D.;  
RT "An HPI-like protein is missing from transcriptionally silent  
RT micronuclei of Tetrahymena.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:13624-13629(1998).  
DR EMBL; AF079405; AAC78328.1;  
DR InterPro; IPR000953; Chromo.  
DR Pfam; PF00385; Chromo. 1.  
DR SMART; SM00298; CHROMO\_2; 1.  
DR PROSITE; PSS0013; CHROMO\_2; 1.  
SQ SEQUENCE 184 AA; 21140 MW; 8B38D646B09CD38F CRC64;  
Query Match  
Best Local Similarity 81.6%; Score 31; DB 5; Length 184;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 KEIHLX 7  
|||||  
Db 115 KEIHLX 121  
RESULT 11  
073155 PRELIMINARY; PRT; 233 AA.  
AC 073155;  
DT 01-FEB-1997 (TREMblrel\_02, Created)  
DT 01-FEB-1997 (TREMblrel\_02, Last sequence update)  
DT 01-OCT-2001 (TREMblrel\_18, Last annotation update)  
DE HYPOTHETICAL 26.1 KDA PROTEIN.  
GN SLR1039.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_Taxid=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hikosawa M., Suglura M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,  
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
DR EMBL; D90904; BAA17181.1;  
DR InterPro; IPR001601; Meth-transf.  
DR InterPro; IPR000051; SAM\_bind.  
DR InterPro; IPR004033; ubiE\_COO5\_methyltransf.  
DR Pfam; PF01209; ubiE\_methyltran; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 233 AA; 26104 MW; 681F5B03AABD288 CRC64;  
Query Match  
Best Local Similarity 81.6%; Score 31; DB 16; Length 233;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 EIHLYX 7  
|:|||||  
Db 208 EIHLYX 213  
RESULT 12  
09FHY9 PRELIMINARY; PRT; 506 AA.  
AC 09FHY9;  
DT 01-MAR-2001 (TREMblrel\_16, Created)  
DT 01-MAR-2001 (TREMblrel\_16, Last sequence update)  
DT 01-JUN-2001 (TREMblrel\_17, Last annotation update)  
DE GTPASE ACTIVATOR PROTEIN OF RAB-LIKE SMALL GTPASES-LIKE PROTEIN.  
OS Arabidopsis thaliana (mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=99397451; PubMed=10470850;  
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.  
RT Sequence features of the regions of 1,011,550 bp covered by seventeen  
RT pl and TAC clones.";  
RL DNA Res. 6:183-195(1999).  
DR EMBL; AB017067; BAB08427.1;  
DR InterPro; IPR000195; RabGAP\_TBC.  
DR Pfam; PF00566; TBC; 1.  
DR SMART; SM00164; TBC; 1.  
SQ SEQUENCE 506 AA; 57954 MW; 3132E977A118C5D5 CRC64;  
Query Match  
Best Local Similarity 81.6%; Score 31; DB 10; Length 506;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KEIHLX 7  
|:|||||  
Db 377 KDIHLX 383  
RESULT 13  
017412 PRELIMINARY; PRT; 591 AA.  
AC 017412;  
DT 01-JAN-1999 (TREMblrel\_09, Created)  
DT 01-DEC-2001 (TREMblrel\_19, Last sequence update)  
DT 01-DEC-2001 (TREMblrel\_19, Last annotation update)  
DE B0001.3 PROTEIN.  
GN B0001.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdiloidea;  
OC Rhabdilitidae; Peloderinae; Caenorhabditis.  
OX NCBI\_Taxid=6239;  
RN [1]



```

RA SEQUENCE FROM N.A.
RA Sims M.A.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none.
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: 269634; CAA93452.2;
SQ SEQUENCE 591 AA; 67490 MW; 0B4BA802581D0393 CRC64;

Query Match
Best Local Similarity 57.1%; Score 31; DB 5; Length 591;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
DB 214 KDVLXQ 220

RESULT 14
O9CNH8 PRELIMINARY; PRT; 634 AA.
AC O9CNH8;
ID 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HYPOTHEICAL PROTEIN PM0453.
GN PM0453.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genome sequence of Pasteurella multocida pm70."
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006081; AK02537.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 634 AA; 72154 MW; BE521565BC5145C8 CRC64;

Query Match
Best Local Similarity 71.4%; Score 31; DB 16; Length 634;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
DB 41 KTVHLXQ 47

RESULT 15
O07549 PRELIMINARY; PRT; 673 AA.
AC O07549;
ID 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHEICAL 76.3 KDA PROTEIN.
GN YHEH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

```

```

RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hajeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones U.,
RA Joris B., Katamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schoefer R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL: Y14080; CAA7444.1;
DR EMBL: Z59109; CAB12811.1;
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmem.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane_1.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 673 AA; 76304 MW; 533DC9970E9B8D5 CRC64;

```

```

Query Match
Best Local Similarity 71.4%; Score 31; DB 16; Length 673;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KEIHLXQ 7
DB 86 KEAHLXQ 92

```

Search completed: August 20, 2002, 11:31:55  
 Job time: 1469 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:16 ; Search time 187.18 seconds  
(without alignments)  
4.154 Million cell updates/sec

Title: US-09-824-286-15  
Perfect score: 34  
Sequence: 1 LQNLVIP 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_032802:\*

1: /SID5/gcgdata/geneseq/genesqp-emb1/AA1980.DAT:\*

2: /SID5/gcgdata/geneseq/genesqp-emb1/AA1981.DAT:\*

3: /SID5/gcgdata/geneseq/genesqp-emb1/AA1982.DAT:\*

4: /SID5/gcgdata/geneseq/genesqp-emb1/AA1983.DAT:\*

5: /SID5/gcgdata/geneseq/genesqp-emb1/AA1984.DAT:\*

6: /SID5/gcgdata/geneseq/genesqp-emb1/AA1985.DAT:\*

7: /SID5/gcgdata/geneseq/genesqp-emb1/AA1986.DAT:\*

8: /SID5/gcgdata/geneseq/genesqp-emb1/AA1987.DAT:\*

9: /SID5/gcgdata/geneseq/genesqp-emb1/AA1988.DAT:\*

10: /SID5/gcgdata/geneseq/genesqp-emb1/AA1989.DAT:\*

11: /SID5/gcgdata/geneseq/genesqp-emb1/AA1990.DAT:\*

12: /SID5/gcgdata/geneseq/genesqp-emb1/AA1991.DAT:\*

13: /SID5/gcgdata/geneseq/genesqp-emb1/AA1992.DAT:\*

14: /SID5/gcgdata/geneseq/genesqp-emb1/AA1993.DAT:\*

15: /SID5/gcgdata/geneseq/genesqp-emb1/AA1994.DAT:\*

16: /SID5/gcgdata/geneseq/genesqp-emb1/AA1995.DAT:\*

17: /SID5/gcgdata/geneseq/genesqp-emb1/AA1996.DAT:\*

18: /SID5/gcgdata/geneseq/genesqp-emb1/AA1997.DAT:\*

19: /SID5/gcgdata/geneseq/genesqp-emb1/AA1998.DAT:\*

20: /SID5/gcgdata/geneseq/genesqp-emb1/AA1999.DAT:\*

21: /SID5/gcgdata/geneseq/genesqp-emb1/AA2000.DAT:\*

22: /SID5/gcgdata/geneseq/genesqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	7	AAW31652	Human cytokine rec
2	34	100.0	230	AAAR47151	IL-2 receptor gamma
3	34	100.0	230	AAAR82934	Interleukin 4 comp
4	34	100.0	252	AAAR47150	IL-2 receptor gamma
5	34	100.0	347	AAAR47149	IL-2 receptor gamma
6	34	100.0	369	AAAR47148	IL-2 receptor gamma
7	34	100.0	369	AAAR59094	Murine IL-2R gamma
8	34	100.0	482	AAW31646	Human cytokine rec
9	34	100.0	691	AAV92202	Fusion polypeptide
10	34	100.0	694	AAV92201	Fusion polypeptide
11	34	100.0	694	AAV92203	Fusion polypeptide

12	31	91.2	1374	19	AAW72225	HSV-2 strain SB5 C
13	31	91.2	1374	19	AAW69753	Herpes simplex vir
14	31	91.2	1384	19	AAW72224	HSV-2 strain SB5 C
15	31	91.2	1396	19	AAW72117	HSV-2 strain SB5 C
16	31	91.2	1396	19	AAW72039	HSV-2 strain SB5 C
17	30	88.2	81	22	ABBI6652	Human nervous syst
18	30	88.2	303	21	AAV77129	Marburg virus stru
19	30	88.2	553	21	AAV97004	S. cerevisiae esse
20	30	88.2	634	22	ABR60004	Drosophila melanog
21	29	85.3	125	22	AAW50109	C glutamicum prote
22	29	85.3	232	21	AAW50179	Arabidopsis thalia
23	29	85.3	246	22	AAU37393	Staphylococcus aur
24	29	85.3	246	22	AAU37447	Staphylococcus aur
25	29	85.3	315	21	AAW50193	Arabidopsis thalia
26	29	85.3	363	22	AAW71682	Bos taurus Interle
27	29	85.3	454	21	AAW50178	Arabidopsis thalia
28	29	85.3	537	21	AAW50192	Arabidopsis thalia
29	29	85.3	623	19	AAW75773	Amino acid sequenc
30	29	85.3	625	22	AAU02044	B. thuringiensis t
31	29	85.3	633	22	AAW39511	Human polypeptide
32	29	85.3	643	22	AAW41297	Human polypeptide
33	28	82.4	54	22	AAW06181	Peptide #4863 enco
34	28	82.4	73	21	AAW58387	Lung cancer associ
35	28	82.4	73	22	AAW74277	Human colon cancer
36	28	82.4	83	22	AAW56710	Human colon cancer
37	28	82.4	84	17	AAW99710	PRTI1799-encoded p
38	28	82.4	86	21	AAW44184	Human cancer assoc
39	28	82.4	103	22	AAW62532	B. melitensis vitru
40	28	82.4	192	13	AAW22953	Promoter-secretion
41	28	82.4	192	14	AAW47660	Amino acid sequenc
42	28	82.4	244	20	AAW37743	Arabidopsis thalia
43	28	82.4	257	21	AAW36350	Arabidopsis thalia
44	28	82.4	263	21	AAW36349	Arabidopsis thalia
45	28	82.4	269	21	AAW36348	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAW31652	standard; Peptide; 7 AA.
AAW31652:	
21-MAY-1998	(first entry)
Human cytokine receptor gc chain epitope.	
Cytokine receptor; gamma common chain; gc chain; human;	
blocking agent; monoclonal antibody; CP-B8; immunological disease;	
myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;	
insulin-dependent diabetes; inflammatory bowel disease;	
sympathetic ophthalmia; uveitis; allergy; asthma; infection;	
graft versus host disease; psoriasis; immunosuppressive; therapy;	
epitope.	
OS	Homo sapiens.
XX	
PN	WO9743416-A1.
XX	
PD	20-NOV-1997.
XX	
PF	09-MAY-1997; 97WO-US07870.
XX	
PR	10-MAY-1996; 96US-0017466.
XX	
PA	(BIOG ) BIOGEN INC.
XX	
PI	Benjamin CD, Burky LC, Hession C, Whitty A;
XX	
DR	WPI; 1998-008885/01.
XX	

PT Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell anergy for  
 PT treatment of immunological diseases  
 XX  
 PS Claim 24; Page 84; 111pp; English.  
 CC This peptide comprises an epitope of the human cytokine receptor  
 CC common gamma (gc) chain (see AAW31646) that is recognised by  
 CC gc blocking agents of the invention. 5 Such epitopes (see  
 CC AAW31650-54) have been identified. The invention provides  
 CC compositions and methods for inhibiting cytokine signalling using  
 CC gc chain blocking agents for the treatment of immunological  
 CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,  
 CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel  
 CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,  
 CC parasitic infection, graft vs. host disease or psoriasis. A  
 CC preferred gc blocking agent is Mab CP.B8 or its Fab fragment (see  
 CC also AAW31647-48).  
 XX  
 SQ Sequence 7 AA;  
 Query Match 100.0%; Score 34; DB 19; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 6,4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LQNLVIP 7  
 Ds 1 lgnlvip 7  
 RESULT 2  
 AAR47151  
 ID AAR47151 standard; Protein: 230 AA.  
 XX AAR47151:  
 AC 13-JUN-1994 (first entry)  
 DT 13-JUN-1994 (first entry)  
 DE IL-2 receptor gamma chain.  
 XX  
 XX Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;  
 KM rheumatoid arthritis; transplant rejection; primer;  
 KW polymerase chain reaction; PCR; amplification.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP578932-A.  
 PD 19-JAN-1994.  
 XX  
 XX 22-APR-1993; 93EP-0106561.  
 PF 22-APR-1993; 93EP-0106561.  
 XX  
 PR 23-APR-1992; 92JP-0104947.  
 PS 23-APR-1992; 92JP-0104947.  
 XX  
 PA (AJIN ) AJINOMOTO KK.  
 PA (SUGA/) SUGAMURA K.  
 XX  
 PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;  
 PI Suzuki M, Takeshita T;  
 XX  
 XX WPI: 1994-017546/03.  
 DR N-PSDB; AA054831.  
 XX  
 XX DNA and protein sequences of IL-2 gamma chain - useful as immune  
 PT regulatory agents for treatment of e.g. rheumatoid arthritis and  
 PT transplant rejection  
 XX  
 PS Disclosure; Page 22-23, 35-36; 50pp; English.  
 XX  
 CC The human IL-2 receptor gamma chain preform (AAR47148), including the  
 CC signal peptide, is encoded by the sequence given in AA054828. The  
 CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble

CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,  
 CC while a soluble form suitable for expression in prokaryotes (AAR47151)  
 CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-  
 CC terminal sequence of IL-2 receptor gamma chain, and are used to  
 CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27  
 CC are used to obtain the protein given in AAR47151.  
 XX  
 SQ Sequence 230 AA;  
 Query Match 100.0%; Score 34; DB 15; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LQNLVIP 7  
 Ds 126 lgnlvip 132  
 RESULT 3  
 AAR82934  
 ID AAR82934 standard; Protein: 230 AA.  
 XX AAR82934:  
 AC 26-FEB-1996 (first entry)  
 DT 26-FEB-1996 (first entry)  
 DE Interleukin 4 component common to the IL-2 receptor gamma chain.  
 XX  
 XX Interleukin-4; IL-4; gamma chain component; immunosuppressants;  
 KW anti-allergy agent; signal transduction inhibitor; atcoimmune;  
 KW disease; anti-inflammatory; anaphylactic shock; bronchial asthma;  
 KW Interleukin-2; IL-2; atopic dermatitis; urticaria.  
 XX  
 OS Homo sapiens.  
 XX  
 PN JP07149662-A.  
 PD 13-JUN-1995.  
 XX  
 XX 07-SEP-1994; 94JP-0213706.  
 PF 07-SEP-1994; 94JP-0213706.  
 XX  
 PR 08-SEP-1993; 93JP-0223574.  
 PS 08-SEP-1993; 93JP-0223574.  
 XX  
 PA (AJIN ) AJINOMOTO KK.  
 PA (SUGA/) SUGAMURA K.  
 XX  
 DR WPI: 1995-243601/32.  
 DR N-PSDB; AAT04952.  
 XX  
 XX Novel interleukin-4 receptor monoclonal antibodies inhibit signal  
 PT transmission - useful as immunosuppressants and anti-allergy agents.  
 PT  
 PS Example 1; Page 9; 11pp; Japanese.  
 XX  
 CC AAT04952 encodes AAR82934 a component of the IL-4 receptor common to  
 CC the IL-2 receptor gamma chain molecule, which was used to generate  
 CC anti-IL-4 receptor monoclonal antibodies (mAbs). The mAbs (IL-4  
 CC signal transduction inhibitors) can be used as immunosuppressants  
 CC and anti-allergy agents, for the treatment of autoimmune and chronic  
 CC inflammatory diseases, e.g. anaphylactic shock, bronchial asthma,  
 CC atopic dermatitis and urticaria.  
 XX  
 SQ Sequence 230 AA;  
 Query Match 100.0%; Score 34; DB 16; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 LQNLVIP 7  
 Ds 127 lgnlvip 133

```

RESULT 4
ID AAR47150 standard; Protein; 252 AA.
XX
AC AAR47150;
XX
DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer;
KW polymerase chain reaction; PCR; amplification; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Sig_peptide
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
PA (AJIN ) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI; 1994-017546/03.
DR N-PSDB; AA054830.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Disclosure; Page 21-22, 34-35; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 252 AA;

```

```

Query Match 100.0%; Score 34; DB 15; Length 252;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 LQNLVIP 7
   |||||
Db 148.Lqnlvip 154

```

```

RESULT 5
ID AAR47149 standard; Protein; 347 AA.
XX
AC AAR47149;
XX

```

```

DT 13-JUN-1994 (first entry)
XX
DE IL-2 receptor gamma chain.
XX
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer;
KW polymerase chain reaction; PCR; amplification.
XX
OS Homo sapiens.
XX
PN EP578932-A.
XX
PD 19-JAN-1994.
XX
PE 22-APR-1993; 93EP-0106561.
XX
PR 23-APR-1992; 92JP-0104947.
XX
PA (AJIN ) AJINOMOTO KK.
PA (SUGA/) SUGAMURA K.
XX
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
PI Suzuki M, Takeshita T;
XX
DR WPI; 1994-017546/03.
DR N-PSDB; AA054829.
XX
PT DNA and protein sequences of IL-2 gamma chain - useful as immune
PT regulatory agents for treatment of e.g. rheumatoid arthritis and
PT transplant rejection
XX
PS Claim 4; Page 41; 50pp; English.
XX
CC The human IL-2 receptor gamma chain preform (AAR47148), including the
CC signal peptide, is encoded by the sequence given in AA054828. The
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,
CC while a soluble form suitable for expression in prokaryotes (AAR47151)
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-
CC terminal sequence of IL-2 receptor gamma chain, and are used to
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27
CC are used to obtain the protein given in AAR47151.
XX
SQ Sequence 347 AA;

```

```

Query Match 100.0%; Score 34; DB 15; Length 347;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 LQNLVIP 7
   |||||
Db 126.Lqnlvip 132

```

```

RESULT 6
ID AAR47148 standard; Protein; 369 AA.
XX
AC AAR47148;
XX

```

```

DT 13-JUN-1994 (first entry)
XX

```

```

DE IL-2 receptor gamma chain.
XX

```

```

KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;
KW rheumatoid arthritis; transplant rejection; primer; PCR;
KW polymerase chain reaction; amplification; ss.
XX

```

```

OS Homo sapiens.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1..22

```

```

FT XX /label= Sig_peptide
PN XX EP578932-A.
XX PD 19-JAN-1994.
XX PF 22-APR-1993; 93EP-0106561.
XX PR 23-APR-1992; 92JP-0104947.
XX PA (AJIN ) AJINOMOTO KK.
XX PI (SUGA/) SUGAMURA K.
XX PI Aaao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;
XX PI Suzuki M, Takeshita T;
XX DR WPI: 1994-017546/03.
XX DR N-PSDB: AAQ54828.
XX PT DNA and protein sequences of IL-2 gamma chain - useful as immune
XX PT regulatory agents for treatment of e.g. rheumatoid arthritis and
XX PT transplant rejection
XX PS Disclosure; Page 16-17, 29-30; 50pp; English.
XX CC The human IL-2 receptor gamma chain preform (AAK47148), including the
XX CC signal peptide, is encoded by the sequence given in AAQ54828. The
XX CC mature protein (AAK47149) is encoded by sequence AAQ54829. A soluble
XX CC form of IL-2 receptor gamma chain (AAK47150) is encoded by AAQ54830,
XX CC while a soluble form suitable for expression in prokaryotes (AAK47151)
XX CC is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-
XX CC terminal sequence of IL-2 receptor gamma chain, and are used to
XX CC isolate IL2 receptor gamma chain receptor cDNA. Primers AAQ54826-27
XX CC are used to obtain the protein given in AAK47151.
XX SQ Sequence 369 AA;

Query Match 100.0%; Score 34; DB 15; Length 369;
Best Local Similarity 100.0%; Pred. NO. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
DB 148 lqnlvip 154

RESULT 7
AAK59094
ID AAK59094 standard; Protein: 369 AA.
XX AC AAK59094;
XX DT 04-MAY-1995 (first entry)
XX DE Murine IL-2R gamma.
XX KW Murine IL2-R gamma; X-linked severe combined immunodeficiency;
XX KW XSCID; Interleukin.
XX OS Mus musculus.
XX FH Key Location/Qualifiers
XX FH Peptide 1..21
XX FT /note= "signal peptide"
XX FT Domain 258..284
XX FT /note= "transmembrane domain"
XX FT Misc-difference 331
XX FT /note= "Corresponding codon CAG"
XX FT Modified-site 71..73
XX FT /label= N-glycosylation_site
XX FT Modified-site 75..77
XX FT /label= N-glycosylation_site

```

```

FT FT Modified-site 84..86
FT FT /label= N-glycosylation_site
FT FT Modified-site 96..98
FT FT /label= N-glycosylation_site
FT FT Modified-site 159..161
FT FT /label= N-glycosylation_site
FT FT Modified-site 255..257
FT FT /label= N-glycosylation_site

XX PN W09420641-A.
XX PD 15-SEP-1994.
XX XX 10-MAR-1994; 94WO-US02891.
XX PF 12-MAR-1993; 93US-0031143.
XX PR 14-SEP-1993; 93US-0121435.
XX XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PA Leonard WJ, McBride WO, Noguchi M;
XX PI WPI: 1994-303046/37.
XX DR N-PSDB: AAQ71977.
XX CC Diagnosis of X-linked severe combined immunodeficiency (XSCID) -
XX CC comprises detecting mutated IL-2R gamma gene, also vectors and
XX CC transgenic animals containing the mutated gene
XX PS Example 1; Fig 7; 98pp; English.
XX CC AAK71977 is the DNA sequence of murine IL-2R gamma AAK59094,
XX CC this was used in the development of a claimed method for the
XX CC diagnosis of X-linked severe combined immunodeficiency (XSCID),
XX CC in female carriers and male sufferers.
XX SQ Sequence 369 AA;

Query Match 100.0%; Score 34; DB 15; Length 369;
Best Local Similarity 100.0%; Pred. NO. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
DB 148 lqnlvip 154

RESULT 8
AAW31646
ID AAW31646 standard; Protein: 482 AA.
XX AC AAW31646;
XX DT 21-MAY-1998 (first entry)
XX DE Human cytokine receptor gc chain-Ig fusion protein.
XX KW Cytokine receptor; gamma common chain; gc chain; human;
XX KW blocking agent; monoclonal antibody; Cp.B8; Immunological disease;
XX KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;
XX KW insulin-dependent diabetes; inflammatory bowel disease;
XX KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;
XX KW graft versus host disease; psoriasis; Immunosuppressive; therapy.
XX OS Chimeric - Homo sapiens.
XX FH Key Location/Qualifiers
XX FH Protein 1..254
XX FT /note= "gc chain N-terminal region"
XX FT Protein 255..482
XX FT /note= "IgG1 constant region"
XX FT Region 255..264

```

FT FT /note="IgG1 hinge region"  
 FT Domain 264...482  
 FT /note="IgG1 CH2 and CH3 constant domains1"  
 XX  
 PN WO9743416-A1.  
 XX  
 XX 20-NOV-1997.  
 PD  
 XX  
 PF 09-MAY-1997; 97WO-US07870.  
 XX  
 PR 10-MAY-1996; 96US-0017466.  
 XX  
 PA (BIOJ ) BIOGEN INC.  
 XX  
 PI Benjamin CD, Burkly LC, Hession C, Whitty A;  
 XX  
 DR WPI: 1998-008885/01.  
 DR N-PSDB: AAT97439.  
 XX  
 PT Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell anergy for  
 PT treatment of immunological diseases  
 XX  
 PS Example 1; Page 79-80; 11pp; English.  
 XX  
 CC This polypeptide comprises a fusion between the N-terminal 254  
 CC amino acids of the human mature cytokine receptor gamma common (gc)  
 CC chain and the hinge region and CH2 and CH3 constant domains of  
 CC human IgG1. The fusion was expressed from clone pLB001 (see  
 CC AAT97439) in COS-7 cells, and used to generate murine anti-human gc  
 CC specific monoclonal antibodies (Mabs), including CP.B8 produced by  
 CC hybridoma ATCC HB 12107. The invention provides compositions and  
 CC methods for inhibiting cytokine signalling using gc chain blocking  
 CC agents for the treatment of immunological diseases such as  
 CC myasthenia gravis, rheumatoid arthritis, lupus, multiple sclerosis,  
 CC insulin-dependent diabetes, inflammatory bowel disease, sympathetic  
 CC ophthalmia, uveitis, allergy, asthma, parasitic infection, graft  
 CC vs. host disease or psoriasis. A preferred gc blocking agent is  
 CC Mab CP.B8 or its Fab fragment (see also AAW31647-48).  
 XX  
 SQ Sequence 482 AA:  
 OY 1 LQNLVIP 7  
 DB 148 lqnlvip 154  
 Query Match 100.0%; Score 34; DB 19; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 9  
 AAY92202  
 ID AAY92202 standard; Protein; 691 AA.  
 XX  
 AC AAY92202;  
 XX  
 DT 01-AUG-2000 (first entry)  
 XX  
 DE Fusion polypeptide 603, IL-4 trap.  
 XX  
 KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;  
 KW cytostatic; immunomodulator; osteopathic.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO200018932-A2.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 22-SEP-1999; 99WO-US22045.

XX  
 PR 25-SEP-1998; 98US-0101858.  
 PR 19-MAY-1999; 99US-0313942.  
 XX  
 PA (REGG-) REGENERON PHARM INC.  
 XX  
 PI Stahl N, Yancopoulos GD;  
 XX  
 DR WPI: 2000-293165/25.  
 DR N-PSDB: AAA09044.  
 XX  
 PT Isolated nucleic acid molecule for treating cytokine-related diseases  
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine  
 PT to form a nonfunctional complex  
 XX  
 PS Example 6; Fig 22A-D; 152pp; English.  
 XX  
 CC This sequence shows fusion polypeptide 603, which is capable of  
 CC binding cytokine IL-4 to form a non-functional complex.  
 CC The invention concerns production of antagonists to any cytokine that  
 CC utilizes an alpha specificity determining component, which when combined  
 CC with the cytokine, binds to a first beta signal transducing component to  
 CC form a non-functional intermediate which then binds to a second beta  
 CC signal transducing component causing beta-receptor dimerization, the  
 CC soluble alpha specificity determining component of the receptor  
 CC (SR-alpha) and the extracellular domain of the first beta signal  
 CC transducing component of the cytokine receptor (beta-1) are combined to  
 CC form heterodimers (sr-alpha:beta-1) that act as antagonist to the  
 CC cytokine by binding the cytokine to form a non-functional complex. The  
 CC receptor components are shared by cytokines such as the CNTF (ciliary  
 CC neurotrophic factor) family of cytokines. The invention provides the  
 CC basis for the development of IL-6 antagonists, as they show that if, in  
 CC the presence of a ligand, a non-functional intermediate complex,  
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor  
 CC component, can be formed, it will effectively block the action of the  
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers  
 CC of the extracellular domains of the alpha specificity determining  
 CC components of their receptors and the extracellular domain of gp130.  
 CC The resultant heterodimers, function as high-affinity traps, rendering  
 CC the cytokine inaccessible to form a signal transducing complex with the  
 CC native membrane-bound forms of their receptor. The nucleic acids and  
 CC polypeptides are useful for treating cytokine-related diseases or  
 CC disorders such as osteoporosis and primary and secondary effects of  
 CC cancer including multiple myeloma or cachexia.  
 XX  
 SQ Sequence 691 AA:  
 OY 1 LQNLVIP 7  
 DB 148 lqnlvip 154  
 Query Match 100.0%; Score 34; DB 21; Length 691;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 10  
 AAY92201  
 ID AAY92201 standard; Protein; 694 AA.  
 XX  
 AC AAY92201;  
 XX  
 DT 01-AUG-2000 (first entry)  
 XX  
 DE Fusion polypeptide 424, IL-4 trap.  
 XX  
 KW IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;  
 KW cytostatic; immunomodulator; osteopathic.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX

PN WO200018932-A2.  
 XX 06-APR-2000.  
 PD  
 XX 22-SEP-1999; 99WO-US22045.  
 PF  
 XX 25-SEP-1998; 98US-0101858.  
 PR 19-MAY-1999; 99US-0313942.  
 XX  
 PA (REG-) REGENERON PHARM INC.  
 XX  
 PI Stahl N, Yancopoulos GD;  
 XX  
 XX WPI: 2000-293165/25.  
 DR N-PSDB; AAA09043.  
 XX  
 XX Isolated nucleic acid molecule for treating cytokine-related diseases  
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine  
 PT to form a nonfunctional complex  
 XX  
 XX Example 6; Fig 21; 152pp; English.  
 PS  
 CC This sequence shows fusion polypeptide 424, which is capable of  
 CC binding cytokine IL-4 to form a non-functional complex.  
 CC The invention concerns production of antagonists to any cytokine that  
 CC utilizes an alpha specificity determining component, which when combined  
 CC with the cytokine, binds to a first beta signal transducing component to  
 CC form a non-functional intermediate which then binds to a second beta  
 CC signal transducing component causing beta-receptor dimerization, the  
 CC soluble alpha specificity determining component of the receptor  
 CC (SR-alpha) and the extracellular domain of the first beta signal  
 CC transducing component of the cytokine receptor (beta-1) are combined to  
 CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the  
 CC cytokine by binding the cytokine to form a non-functional complex. The  
 CC receptor components are shared by cytokines such as the CNTF (ciliary  
 CC neurotrophic factor) family of cytokines. The invention provides the  
 CC basis for the development of IL-6 antagonists, as they show that if, in  
 CC the presence of a ligand, a non-functional intermediate complex,  
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor  
 CC component, can be formed, it will effectively block the action of the  
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers  
 CC of the extracellular domains of the alpha specificity determining  
 CC components of their receptors and the extracellular domain of gp130.  
 CC The resultant heterodimers, function as high-affinity traps, rendering  
 CC the cytokine inaccessible to form a signal transducing complex with the  
 CC native membrane-bound forms of their receptor. The nucleic acids and  
 CC polypeptides are useful for treating cytokine-related diseases or  
 CC disorders such as osteoporosis and primary and secondary effects of  
 CC cancer including multiple myeloma or cachexia.  
 CC  
 XX  
 SO Sequence 694 AA:  
 Query Match 100.0%; Score 34; DB 21; Length 694;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 LQNLVIP 7  
 Db 148 lqnlvip 154  
 RESULT 11  
 ID AAY92203 standard; Protein: 694 AA.  
 XX  
 AC AAY92203;  
 XX  
 DT 01-AUG-2000 (first entry)  
 XX  
 DE Fusion polypeptide 622, IL-4 trap.  
 XX  
 XX IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;  
 KM

KW cyostatic; immunomodulator; osteopathic.  
 XX Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO200018932-A2.  
 PD  
 XX 06-APR-2000.  
 PF  
 XX 22-SEP-1999; 99WO-US22045.  
 PR 25-SEP-1998; 98US-0101858.  
 PR 19-MAY-1999; 99US-0313942.  
 XX  
 PA (REG-) REGENERON PHARM INC.  
 XX  
 PI Stahl N, Yancopoulos GD;  
 XX  
 XX WPI: 2000-293165/25.  
 DR N-PSDB; AAA09045.  
 XX  
 XX Isolated nucleic acid molecule for treating cytokine-related diseases  
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine  
 PT to form a nonfunctional complex  
 XX  
 XX Example 6; Fig 23A-D; 152pp; English.  
 PS  
 CC This sequence shows fusion polypeptide 622, which is capable of  
 CC binding cytokine IL-4 to form a non-functional complex.  
 CC The invention concerns production of antagonists to any cytokine that  
 CC utilizes an alpha specificity determining component, which when combined  
 CC with the cytokine, binds to a first beta signal transducing component to  
 CC form a non-functional intermediate which then binds to a second beta  
 CC signal transducing component causing beta-receptor dimerization, the  
 CC soluble alpha specificity determining component of the receptor  
 CC (SR-alpha) and the extracellular domain of the first beta signal  
 CC transducing component of the cytokine receptor (beta-1) are combined to  
 CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the  
 CC cytokine by binding the cytokine to form a non-functional complex. The  
 CC receptor components are shared by cytokines such as the CNTF (ciliary  
 CC neurotrophic factor) family of cytokines. The invention provides the  
 CC basis for the development of IL-6 antagonists, as they show that if, in  
 CC the presence of a ligand, a non-functional intermediate complex,  
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor  
 CC component, can be formed, it will effectively block the action of the  
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers  
 CC of the extracellular domains of the alpha specificity determining  
 CC components of their receptors and the extracellular domain of gp130.  
 CC The resultant heterodimers, function as high-affinity traps, rendering  
 CC the cytokine inaccessible to form a signal transducing complex with the  
 CC native membrane-bound forms of their receptor. The nucleic acids and  
 CC polypeptides are useful for treating cytokine-related diseases or  
 CC disorders such as osteoporosis and primary and secondary effects of  
 CC cancer including multiple myeloma or cachexia.  
 CC  
 XX  
 SO Sequence 694 AA:  
 Query Match 100.0%; Score 34; DB 21; Length 694;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 LQNLVIP 7  
 Db 148 lqnlvip 154  
 RESULT 12  
 ID AAW72225 standard; Protein: 1374 AA.  
 XX  
 AC AAW72225;  
 XX



DT 13-JAN-1999 (first entry)  
XX  
DE HSV-2 strain SB5 Contig ID 15 ORF#20b protein.  
XX  
XX HSV-2 strain SB5: immunological response induction; therapy;  
KM antiviral identification; viral protein inhibitor.  
XX  
OS Herpes simplex virus type 2.  
XX  
PN WO9820016-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 31-OCT-1997; 97WO-US20016.  
XX  
PR 09-JUN-1997; 97US-0049018.  
PR 04-NOV-1996; 96US-0030279.  
XX  
XX (SMIR ) SMITHKLINE BEECHAM CORP.  
PA Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;  
PI Esser KM, Leary JY;  
XX  
DR WPI: 1998-286847/25.  
DR N-PSDB; AAV62176.  
XX  
PT Herpes simplex virus type-2 sequences - useful in, e.g., prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal  
XX  
PS Claim 10; Page 141-142; 748bp; English.  
XX  
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.  
CC Based on homology, this sequence is a major capsid protein.  
CC The proteins can be used for the treatment or prevention of disease, to  
CC induce an immunological response in a mammal or to identify inhibitors,  
CC activators or novel antivirals. Antagonists of the proteins can be used  
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
CC it can also be used to induce an immunological response in a mammal.  
XX  
SQ Sequence 1374 AA;  
OY  
Query Match 91.2%; Score 31; DB 19; Length 1374;  
Best Local Similarity 71.4%; Pred. No. 5e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 LONLVIP 7  
DB 839 lqnmvvp 845  
RESULT 13  
AAW69753  
ID AAW69753 standard; protein; 1374 AA.  
XX  
AC AAW69753;  
XX  
DT 02-NOV-1998 (first entry)  
XX  
DE Herpes simplex virus-1 VP5 protein.  
XX  
KM Herpes simplex virus-1; HSV-1; UL15; VP5; UL19; antiviral agent;  
KM identification.  
XX  
OS Herpes simplex virus type 1.  
XX  
PN EP860700-A2.  
XX  
PD 26-AUG-1998.  
XX

PF 16-FEB-1998; 98EP-0301123.  
XX  
XX 21-FEB-1997; 97US-0038914.  
XX  
XX (SMIR ) SMITHKLINE BEECHAM CORP.  
PA Del Vecchio A;  
PI  
XX  
DR WPI: 1998-439489/38.  
XX  
PT Screening assays for antiviral agents - based on inhibition of  
PT herpes simplex UL15 and VP5 interaction or function  
XX  
XX Claim 10; Page 11-16; 18pp; English.  
XX  
CC Methods have been developed for: (1) identifying an antiviral agent  
CC useful in treating infection with herpes viruses, comprising: (a)  
CC providing UL15 of herpes simplex virus-1 (HSV-1) or a functional  
CC derivative or homologue thereof; (b) permitting UL15 or its derivative  
CC or homologue to come into association with a test compound; (c)  
CC providing VP5 of HSV-1 or a functional derivative or homologue thereof;  
CC (d) screening for inhibition of the interaction between UL15 or its  
CC derivative or homologue and VP5 or its derivative or homologue and VP5  
CC or its derivative or homologue, where such inhibition is indicative that  
CC the test compound is an antiviral agent; (2) identifying an antiviral  
CC agent useful in treating infection with herpes viruses, comprising: (a)  
CC providing VP5 of herpes simplex virus-1 (HSV-1) or a functional  
CC derivative or homologue thereof; (b) permitting VP5 or its derivative  
CC or homologue to come into association with a test compound; (c)  
CC providing UL15 of HSV-1 or a functional derivative or homologue  
CC thereof; (d) screening for inhibition of the interaction between the  
CC VP5 or its derivative or homologue and UL15 or its derivative or  
CC homologue, where such inhibition is indicative that the test compound  
CC is an antiviral agent. The present sequence represents the HSV-1 VP5  
CC (or UL15) protein.  
XX  
SQ Sequence 1374 AA;  
OY  
Query Match 91.2%; Score 31; DB 19; Length 1374;  
Best Local Similarity 71.4%; Pred. No. 5e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 LONLVIP 7  
DB 839 lqnmvvp 845  
RESULT 14  
AAW72224  
ID AAW72224 standard; protein; 1384 AA.  
XX  
AC AAW72224;  
XX  
DT 13-JAN-1999 (first entry)  
XX  
DE HSV-2 strain SB5 Contig ID 15 ORF#20a protein.  
XX  
KM HSV-2 strain SB5; immunological response induction; therapy;  
KM antiviral identification; viral protein inhibitor.  
XX  
OS Herpes simplex virus type 2.  
XX  
PN WO9820016-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 31-OCT-1997; 97WO-US20016.  
XX  
PR 09-JUN-1997; 97US-0049018.  
PR 04-NOV-1996; 96US-0030279.  
XX  
XX (SMIR ) SMITHKLINE BEECHAM CORP.  
PA

XX Chan JY, Debrowski-Amara CE, Delvecchio AM, Dillon SB;  
PI Esser KM, Leary JY;  
XX  
DR WPI: 1998-286847/25.  
DR N-PSDB: AAV62176.  
XX  
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal  
XX  
PS Claim 10: Page 140-141: 748bp; English.  
XX  
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.  
CC Based on homology, this sequence is a major capsid protein.  
CC The proteins can be used for the treatment or prevention of disease, to  
CC induce an immunological response in a mammal or to identify inhibitors,  
CC activators or novel antivirals. Antagonists of the proteins can be used  
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
CC it can also be used to induce an immunological response in a mammal.  
XX  
SQ Sequence 1384 AA;

Query Match 91.28; Score 31; DB 19; Length 1384;  
Best Local Similarity 71.4%; Pred. NO. 5.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 LONLVIP 7  
|||:|:|  
Db 849 lqnmvvp 855

## RESULT 15

AAW72117  
ID AAW72117 standard; Protein; 1396 AA.

XX  
AC AAW72117;

XX  
DT 18-DEC-1998 (first entry)

XX  
DE HSV-2 strain SB5 Contig ID 15 ORF#4 protein.

XX  
KM HSV-2 strain SB5; immunological response induction; therapy;  
KW antiviral identification; viral protein inhibitor.

XX  
OS Herpes simplex virus type 2.

XX  
PN W09820016-A1.

XX  
PD 14-MAY-1998.

XX  
PF 31-OCT-1997; 97WO-US20016.

XX  
PR 09-JUN-1997; 97US-0049018.

XX  
PR 04-NOV-1996; 96US-0030279.

XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.

XX  
PI Chan JY, Debrowski-Amara CE, Delvecchio AM, Dillon SB;  
PI Esser KM, Leary JY;

XX  
DR WPI: 1998-286847/25.

XX  
DR N-PSDB: AAV62159.

XX  
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
PT and treatment of infection or inducing immunological response in  
PT mammal

XX  
PS Claim 10: Page 88-89: 748bp; English.

CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
CC sequence of the invention. This sequence was isolated from a HSV-2 strain  
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.  
CC Based on homology, this sequence is a major capsid protein.  
CC The proteins can be used for the treatment or prevention of disease, to  
CC induce an immunological response in a mammal or to identify inhibitors,  
CC activators or novel antivirals. Antagonists of the proteins can be used  
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing  
CC it can also be used to induce an immunological response in a mammal.  
XX  
SQ Sequence 1396 AA;

Query Match 91.28; Score 31; DB 19; Length 1396;  
Best Local Similarity 71.4%; Pred. NO. 5.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
OY 1 LONLVIP 7  
|||:|:|  
Db 861 lqnmvvp 867

Search completed: August 20, 2002, 11:07:17  
Job time: 6102 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:36 ; Search time 100.84 Seconds  
(without alignments)  
6.670 Million cell updates/sec

Title: US-09-824-286-15

Perfect score: 34

Sequence: 1 LQNMVIP 7

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.71.\*  
2: pir1.\*  
3: pir2.\*  
4: pir3.\*  
5: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	369	2	149280
2	34	100.0	369	2	A42565
3	34	100.0	373	2	A55718
4	33	97.1	496	2	E70142
5	31	91.2	216	2	AE2548
6	31	91.2	1374	1	VCBE17
7	30	88.2	267	2	SG2488
8	30	88.2	303	2	S32583
9	30	88.2	346	2	T14049
10	30	88.2	361	1	QOCVPT
11	30	88.2	513	2	E97393
12	30	88.2	513	2	AG2611
13	30	88.2	553	2	S53080
14	30	88.2	689	2	T19999
15	30	88.2	1028	2	B24785
16	29	85.3	171	2	T25696
17	29	85.3	184	2	T51655
18	29	85.3	217	2	T14621
19	29	85.3	246	2	E89884
20	29	85.3	261	2	T46138
21	29	85.3	279	2	AC2165
22	29	85.3	345	2	CE4489
23	29	85.3	346	2	T17105
24	29	85.3	346	2	T13835
25	29	85.3	346	2	T14056
26	29	85.3	379	2	S55845
27	29	85.3	513	2	T27355
28	29	85.3	535	2	S44827
29	29	85.3	537	2	T04745

#### ALIGNMENTS

```

RESULT 1
149280
Interleukin-2 receptor gamma chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Feb-1997 #sequence-revision 27-Feb-1997 #extl-change 20-Jun-2000
C:Accession: I49280; A47514; JN0592; S37582; I53398
R:Caio, X.; Kozak, C.A.; Liu, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 8464-8468, 1993
A:Title: Characterization of cDNAs encoding the murine interleukin 2 receptor (IL-2R)
A:Reference number: A47514; MUID:93391374
A:Accession: I49280
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <CAO>
A:Cross-references: EMBL:U21795; NID:g727349; PIDN:AAA64279.1; PID:g727350
A:Accession: A47514
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-369 <RE2>
A:Cross-references: GB:L20048; NID:g404067; PIDN:AAA39286.1; PID:g404068
R:Kumaki, S.; Kondo, M.; Takeshita, T.; Asao, H.; Nakamura, M.; Sugamura, K.
Biochem. Biophys. Res. Commun. 193, 356-363, 1993
A:Title: Cloning of the mouse interleukin 2 receptor gamma chain: Demonstration of fu
A:Reference number: JN0592; MUID:93277575
A:Accession: JN0592
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-369 <KUN>
A:Cross-references: DDBJ:D13565; NID:g303684; PIDN:BAA02760.1; PID:g303685
R:Kobayashi, N.; Nakagawa, S.; Minami, Y.; Taniguchi, T.; Kono, T.
Gene 130, 303-304, 1993
A:Title: Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor gamma.
A:Reference number: JN0775; MUID:93366191
A:Accession: JN0775
A:Molecule type: mRNA
A:Residues: 1-369 <KOB>
A:Cross-references: GB:D13821; NID:g436045; PIDN:BAA02974.1; PID:g436046
R:Chiu, R.K.; Dougherty, G.J.
submitted to the EMBL Data Library, October 1993
A:Description: Regulation of CD44-mediated cellular adhesion by the IL-2 R gamma cha
A:Reference number: S37582
A:Accession: S37582
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-350, S', 352-366, S', 368-369 <CHI>
A:Cross-references: EMBL:X75337
R:Disanto, J.P.; Certain, S.; Wilson, A.; MacDonald, H.R.; Avner, P.; Fischer, A.; de
Eur. J. Immunol. 24, 3014-3018, 1994
A:Title: The murine interleukin-2 receptor gamma chain gene: organization, chromosoma
A:Reference number: I53398; MUID:95104285
A:Accession: I53398
A:Status: preliminary; translated from GB/EMBL/DBJ

```

A:Molecule type: DNA  
 A:Residues: 1-369 <RES>  
 A:Cross-references: GB:S75852; NID:9861554; PIDN:AA32904.1; PID:9861555  
 C:Genetics:  
 A:Gene: IL-2Rgamma  
 A:Introns: 39/1; 90/2; 152/1; 199/3; 254/1; 286/2; 308/3  
 C:Complex: The high affinity receptor is a heterotrimer of alpha (see PIR:UHMS2), beta  
 eptors.  
 C:Function:  
 A:Description: receptor for interleukin-2  
 A:Pathway: Interleukin-2 stimulated growth and differentiation of T cells, B cells, NK  
 C:Superfamily: Interleukin-2 receptor gamma chain  
 C:Keywords: cytokine receptor; glycoprotein; transmembrane protein  
 F:1-22/DNA: signal sequence #status predicted <SIG>  
 F:23-369/Product: interleukin-2 receptor gamma chain #status predicted <MAT>  
 F:256-284/DNA: transmembrane #status predicted <TM>  
 F:1/1,5,84,96,159,164,306/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 34; DB 2; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 Db 148 LQNLVIP 154

RESULT 2  
 A42565  
 Interleukin-2 receptor gamma chain - human  
 C:Species: Homo sapiens (man)  
 C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
 C:Accession: A42565; A46591; F54332  
 R:Takeshita, T.; Asao, H.; Ohtani, K.; Ishii, N.; Kumaki, S.; Tanaka, N.; Munakata, H.;  
 Science 257, 379-382, 1992  
 A:Title: Cloning of the gamma chain of the human IL-2 receptor.  
 A:Reference number: A42565; MWID:92335883  
 A:Accession: A42565  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid; protein  
 A:Residues: 1-369 <PAK>  
 A:Cross-references: GB:D11086; NID:9303611; PIDN:BA401857.1; PID:9319890  
 A:Experimental source: MOLT beta lymphoid cells  
 A:Note: sequence extracted from NCBI backbone (NCIIP:109167)  
 R:Moguchi, M.; Adelstein, S.; Cao, X.; Leonard, W.J.  
 J. Biol. Chem. 268, 13601-13608, 1993  
 A:Title: Characterization of the human interleukin-2 receptor gamma chain gene.  
 A:Reference number: A46591; MWID:93293887  
 A:Accession: A46591  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-369 <RES>  
 A:Cross-references: GB:L12183; NID:9307056; PIDN:AA59145.1; PID:9307058  
 R:Puck, J.M.; Deschenes, S.M.; Porter, J.C.; Dutra, A.S.; Brown, C.J.; Willard, H.F.; He  
 Hum. Mol. Genet. 2, 1099-1104, 1993  
 A:Title: The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-link  
 A:Reference number: F54332; MWID:94004847  
 A:Accession: F54332  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-369 <RE2>  
 A:Cross-references: GB:L19546; NID:9349631; PIDN:AA37524.1; PID:9349632  
 C:Genetics:  
 A:Gene: GDB:IL2RG; SCIDX1; IMD4  
 A:Cross-references: GDB:134807; OMIM:308380  
 A:Map position: Xq13.1-Xq13.1  
 A:Introns: 39/1; 90/2; 152/1; 198/3; 253/1; 285/2; 308/3  
 A:Note: defects are associated with an X-linked form of severe combined immunodeficiency  
 C:Superfamily: Interleukin-2 receptor gamma chain  
 C:Keywords: cytokine receptor; duplication; immunodeficiency; severe combined immunodeficiency

Query Match 100.0%; Score 34; DB 2; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 Db 148 LQNLVIP 154

RESULT 3  
 A55718  
 Interleukin-2 receptor gamma chain precursor - dog  
 C:Species: Canis lupus familiaris (dog)  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 23-Jul-1999  
 C:Accession: A55718  
 R:Henlhorn, P.S.; Somberg, R.L.; Flimant, V.M.; Puck, J.M.; Patterson, D.F.; Felsburg  
 Genomics 23, 69-74, 1994  
 A:Title: IL-2Rgamma gene microdeletion demonstrates that canine X-linked severe combi  
 A:Reference number: A55718; MWID:9510114  
 A:Accession: A55718  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-373 <HEN>  
 A:Cross-references: GB:U04361; NID:9517411; PIDN:AA48403.1; PID:9517412  
 C:Superfamily: Interleukin-2 receptor gamma chain  
 C:Keywords: cytokine receptor; duplication

Query Match 100.0%; Score 34; DB 2; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 Db 148 LQNLVIP 154

RESULT 4  
 E70142  
 Glu-tRNA amidotransferase, subunit A (gluA) homolog - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 22-Jun-1999  
 C:Accession: E70142  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
 son, D.; Peterson, J.; Kervatage, A.R.; Quekeshub, J.; Salzberg, S.; Hanson, M.; Vu  
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MWID:98065943  
 A:Accession: E70142  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-496 <RLE>  
 A:Cross-references: GB:AE001140; GB:AE000783; NID:92688233; PIDN:AA66715.1; PID:9268  
 A:Experimental source: strain B31  
 C:Superfamily: indoleacetamide hydrolase

Query Match 97.1%; Score 33; DB 2; Length 496;  
 Best Local Similarity 85.7%; Pred. No. 12;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 Db 395 LQNLVIP 401

RESULT 5  
 AE2348  
 Hypothetical protein a117672 [imported] - Anabaena sp. (strain PCC 7120) plasmid PCC7  
 C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C/Accession: AE2548  
R/Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iritaghi,  
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A>Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium *Ana*  
A/Reference number: AB1807; MUID:21595285; PMID:11759840  
A/Accession: AE2548  
A>Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-216 <KUR>  
A/Cross-references: GB:AP003602; PIDN:BAW7315.1; PID:g17134757; GSPDB:GN00181  
A/Experimental source: strain PCC 7120  
A/Genetics:  
A/Gene: all7672  
A/Genome: plasmid

Query Match 91.2%; Score 31; DB 2; Length 216;  
Best Local Similarity 71.4%; Pred. No. 14;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7  
|||||  
Db 64 LONLVIP 70

RESULT 6  
VCBE1  
major capsid protein - human herpesvirus 1 (strain 17)  
C/Species: human herpesvirus 1  
C/Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 16-Jun-2000  
C/Accession: A27239; A30084  
R/Davison, A.J.; Scott, J.E.  
J. Gen. Virol. 67, 2279-2286, 1986  
A>Title: DNA sequence of the major capsid protein gene of herpes simplex virus type 1.  
A/Reference number: A27239; MUID:87010565  
A/Accession: A27239  
A/Molecule type: DNA  
A/Residues: 1-1374 <DNA>  
A/Cross-references: GB:X04467; NID:g59857; PIDN:CAA28154.1; PID:g59859  
R/McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perz  
J. Gen. Virol. 69, 1531-1574, 1988  
A>Title: The complete DNA sequence of the long unique region in the genome of herpes sim  
A/Reference number: A30083; MUID:88274327  
A/Accession: A30084  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-1374 <MCG>  
A/Cross-references: GB:X14112; NID:g1944536; PIDN:CAA32332.1; PID:g59519; GB:ID00317  
A/Genetics:  
A/Gene: UL19  
C/Superfamily: varicella-zoster virus major capsid protein  
C/Keywords: capsid protein

Query Match 91.2%; Score 31; DB 1; Length 1374;  
Best Local Similarity 71.4%; Pred. No. 11e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7  
|||||  
Db 839 LONLVIP 845

RESULT 7  
S62488  
hypothetical protein SPAC4G8.11c - fission yeast (*Schizosaccharomyces pombe*)  
C/Species: Schizosaccharomyces pombe  
C/Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Jan-2000  
C/Accession: T36856; S62488  
R/Baddock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, October 1995

A/Reference number: 221745  
A/Accession: T36856  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-267 <BA2>  
A/Cross-references: EMBL:256276; NID:g1022345; PIDN:CAA91212.1; PID:g1022356; GSPDB:G  
A/Experimental source: strain 972b-; cosmid c468  
A/Genetics:  
A/Gene: SPAC4G8.11c  
A/Map position: 1L  
A/Introns: 114/3; 203/3

Query Match 88.2%; Score 30; DB 2; Length 267;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7  
|||||  
Db 6 LONLVIP 12

RESULT 8  
S32583  
structural protein VP40 - Marburg virus (strain Popp)  
C/Species: Marburg virus  
A/Variety: strain Popp  
C/Date: 10-Dec-1993 #sequence\_revision 08-Sep-1995 #text\_change 08-Oct-1999  
C/Accession: S32583; S44051; S28569  
R/Bukreyev, A.A.; Volchkov, V.E.; Bilnov, V.M.; Netesov, S.V.  
FEBS Lett. 322, 41-46, 1993  
A>Title: The VP35 and VP40 proteins of filoviruses. Homology between Marburg and Ebola  
A/Reference number: S32582; MUID:93245956  
A/Accession: S32583  
A/Molecule type: mRNA  
A/Residues: 1-303 <BUK>  
A/Cross-references: EMBL:X64406; NID:g60631; PIDN:CAA45749.1; PID:g60633  
A/Experimental source: strain Popp  
R/Bukreyev, A.A.; Volchkov, V.E.; Bilnov, V.M.; Dryga, S.A.; Netesov, S.V.  
submitted to the EMBL Data Library, January 1994  
A/Description: Full-length nucleotide sequence of Marburg virus Popp strain: The comp  
A/Reference number: S44049  
A/Accession: S44051  
A/Molecule type: genomic RNA  
A/Residues: 1-303 <BUV>  
A/Cross-references: EMBL:229337; NID:g450908; PIDN:CAA82538.1; PID:g450911  
A/Experimental source: strain Popp

Query Match 88.2%; Score 30; DB 2; Length 303;  
Best Local Similarity 71.4%; Pred. No. 36;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7  
|||||  
Db 146 LONLVIP 152

RESULT 9  
T14049  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - *Lamna nasus* mitochondrion (fra  
C/Species: mitochondrion *Lamna nasus*  
A/Note: porbeagle shark  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-Jan-2000  
C/Accession: T14049  
R/Naylor, G.D.P.; Martin, A.P.; Mattison, E.G.; Brown, W.M.  
submitted to the EMBL Data Library, February 1997  
A/Description: Interrelationships of Lamniform Sharks: Testing Phylogenetic Hypotheses  
A/Reference number: Z17808  
A/Accession: T14049  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-346 <NAV>

A:Cross-references: EMBL:U091427; NID:g2257811; PID:g2257812; PIDN:AAB63136.1  
 A:Genetics:  
 A:Genome: mitochondrion  
 A>Note: NADH2  
 A:Superfamily: NADH dehydrogenase (ubiquinone) chain 2  
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 88.2%; Score 30; DB 2; Length 346;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 QNLVIP 7  
 |||||  
 Db 273 QNLVIP 278

# RESULT 10

OCVPT

ALI protein - potato yellow mosaic virus (isolate Venezuela)

C:Species: potato yellow mosaic virus

C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jun-2000

C:Accession: J00364

R:Cutts, R.H.A.; Coffin, R.S.; Roberts, E.J.F.; Hamilton, W.D.O.

J. Gen. Virol. 72, 1515-1520, 1991

A:Title: The nucleotide sequence of the infectious cloned DNA components of potato yellow

A:Reference number: J00362; MUID:91311403

A:Accession: J00364

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1361 <CON>

A:Cross-references: GB:D00940; NID:g222458; PIDN:BAA00782.1; PID:g222459

C:Genetics:

A:Map position: segment A

C:Superfamily: tomato golden mosaic virus ALI protein

Query Match 88.2%; Score 30; DB 1; Length 361;  
 Best Local Similarity 85.7%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QNLVIP 7  
 |||||  
 Db 32 QNLVIP 38

# RESULT 11

E97393

hypothetical protein AGR\_C\_491 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: E97393

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: E97393

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86102.1; PID:g1515183; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C\_491

A:Map position: circular chromosome

Query Match 88.2%; Score 30; DB 2; Length 513;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QNLVIP 7  
 |||||

Db 187 QNLVIP 193

# RESULT 12

AG2611

conserved hypothetical protein Atu0287 [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AG2611

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCl

er, K.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AG2611

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <KUR>

A:Cross-references: GB:AE008688; PIDN:AL41309.1; PID:g17738619; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu0287

A:Map position: circular chromosome

Query Match 88.2%; Score 30; DB 2; Length 513;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 QNLVIP 7  
 |||||  
 Db 187 QNLVIP 193

# RESULT 13

S53080

hypothetical protein YMR258c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR920.12c

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 04-Mar-2000

C:Accession: S53080

R:Hunt, S.; Bowman, S.

submitted to the EMBL Data Library, March 1995

A:Reference number: S53069

A:Accession: S53080

A:Molecule type: DNA

A:Residues: 1-553 <HUN>

A:Cross-references: EMBL:Z48639; NID:g732924; PID:g732936; GSPDB:GN00013; MIPS:YMR258

C:Genetics:

A:Gene: MIPS:YMR258c

A:Map position: 13R

C:Superfamily: Saccharomyces cerevisiae hypothetical protein YMR258c

Query Match 88.2%; Score 30; DB 2; Length 553;  
 Best Local Similarity 71.4%; Pred. No. 71;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 QNLVIP 7  
 |||||  
 Db 192 QNLVIP 198

# RESULT 14

T19999

hypothetical protein C47D12.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T19999

R:Gajadaty, S.

submitted to the EMBL Data Library, March 1996

A;Reference number: Z19209

A;Accession: T19999

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-689 <WILL>

A;Cross-references: EMBL:Z69902; PIDN:CAA93767.1; GSPDB:GN00020; CESP:C47D12.2

A;Experimental source: clone C47D12

C;Genetics:

A;Gene: CESP:C47D12.2

A;Map position: 2

A;Introns: 48/2; 97/2; 167/1; 196/2; 277/1; 439/3; 478/3; 542/3

#### Query Match

88.2%; Score 30; DB 2; Length 689;

Best Local Similarity 71.4%; Pred. No. 91;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNLVIP 7

Db 392 QNLVIP 398

#### RESULT 15

B24785

hypothetical protein 1028 - slime mold (Dictyostelium discoideum) transposon DIRS-1 (fr

C;Species: Dictyostelium discoideum

C;Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 18-Jun-1993

C;Accession: B24785

R;Capello, J.; Handelsman, K.; Lodish, H.F.

Cell 43, 105-115, 1985

A;Title: Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted ter

A;Reference number: A94654; MUID:86079481

A;Accession: B24785

A;Molecule type: DNA

A;Residues: 1-1028 <CAP>

#### Query Match

88.2%; Score 30; DB 2; Length 1028;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QNLVIP 7

Db 447 QNLVIP 452

Search completed: August 20, 2002, 11:10:37  
Job time: 5667 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:01 ; Search time 55.29 Seconds

(Without alignments)  
4.902 Million cell updates/sec

Title: US-09-824-286-15  
Perfect score: 34  
Sequence: 1 LQNLVIP 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	369	1 CYRG_HUMAN	P31785 homo sapien
2	34	100.0	369	1 CYRG_MOUSE	P34902 mus musculu
3	34	100.0	373	1 CYRG_CANFA	P40321 canis famli
4	33	97.1	496	1 GATRA_BORBU	O51317 borrelia bu
5	31	91.2	1374	1 VCAP_HSV11	P06491 herpes simp
6	30	88.2	267	1 VADR_SCHPO	O09836 schizosacch
7	30	88.2	283	1 VP40_MABVM	P35260 marburg vir
8	30	88.2	303	1 VP40_MABVP	O03040 marburg vir
9	30	88.2	361	1 VAL1_PYMVP	P27258 potato yell
10	30	88.2	553	1 YMB8_YEAST	O04847 saccharomyc
11	29	85.3	379	1 CYRG_BOVIN	O05118 bos taurus
12	29	85.3	379	1 YLEU_DEBOC	P48012 debaryomyce
13	29	85.3	535	1 YMA2_CAEEL	P34447 caenorhabdi
14	29	85.3	622	1 ALIB_STRPN	O45743 bacillus th
15	29	85.3	652	1 ALIB_STRPN	O51933 streptococc
16	29	85.3	662	1 CYOB_BUCAI	P57543 buchera ap
17	28	82.4	325	1 V653_STRPY	O5a0r7 streptococc
18	28	82.4	342	1 HRCAL_LEPIN	O51867 leptospira
19	28	82.4	359	1 RFL_CHLMO	O5P116 chlamydia m
20	28	82.4	359	1 RFL_CHLTR	O84026 chlamydia t
21	28	82.4	441	1 YXEK_BACSU	O84026 bacillus su
22	28	82.4	663	1 CYOB_ECOLI	P18401 escherichia
23	28	82.4	666	1 CO14_BRAJA	P98057 bradyrhizob
24	28	82.4	732	1 ADD1_CAEEL	O909X0 caenorhabdi
25	28	82.4	1228	1 ECM_HUMAN	O13201 homo sapien
26	28	82.4	1517	1 YD22_SCHPO	O10250 schizosacch
27	28	82.4	1597	1 RRL1_YEAST	P53552 saccharomyc
28	28	82.4	1950	1 UBRI_YEAST	P19812 saccharomyc
29	27	79.4	221	1 V700_RICPR	O9cmm4 rickettsia
30	27	79.4	268	1 NIKE_ECOLI	P33594 escherichia
31	27	79.4	275	1 NIFH_METWP	O50218 methanococc
32	27	79.4	296	1 VP23_HSV6U	P24436 human herpe
33	27	79.4	405	1 S3AE_BACSU	P49782 bacillus su

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	369 AA.
CYRG_HUMAN				
ID	CYRG_HUMAN			
AC	P31785:			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cytokine receptor common gamma chain precursor (Gamma-C) (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64) (CD132 antigen).			
GN	IL2RG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RA	MEDLINE=92335883; PubMed=1631559;			
RA	Takeshita T., Asao H., Ohtani K., Ishii N., Kumaki S., Tanaka N.,			
RA	Munakata H., Nakamura M., Sugamura K.;			
RT	"Cloning of the gamma chain of the human IL-2 receptor.";			
RL	Science 257:379-382(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RC	MEDLINE=93293887; PubMed=8514792;			
RA	Noguchi M., Adelstein S., Cao X., Leonard W.J.;			
RT	"Characterization of the human Interleukin-2 receptor gamma chain			
RT	gene.";			
RL	J. Biol. Chem. 268:13601-13608(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS ASP-114 AND ASN-153.			
RA	MEDLINE=94004847; PubMed=8401490;			
RA	Puck J.M., Deschenes S.M., Porter J.C., Dutra A.S., Brown C.J.,			
RA	Willard H., Henthorn P.S.;			
RT	"The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated			
RT	in X-linked severe combined immunodeficiency, SCIDX1.";			
RL	Hum. Mol. Genet. 2:1099-1104(1993).			
RN	[4]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RA	MEDLINE=94090315; PubMed=8266076;			
RA	Kondo M., Takeshita T., Ishii N., Nakamura M., Watanabe S.,			
RA	Arai K.-I., Sugamura K.;			
RT	"Sharing of the interleukin-2 (IL-2) receptor gamma chain between			
RT	receptors for IL-2 and IL-4.";			
RL	Science 262:1874-1877(1993).			
RN	[5]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RA	MEDLINE=94090317; PubMed=8266078;			
RA	Russell S.M., Kkegan A.D., Harada N., Nakamura Y., Noguchi M.,			
RA	Leiland P., Friedmann M.C., Miyajima A., Puri R.K., Paul W.E.,			
RA	Leonard W.J.;			
RT	"Interleukin-2 receptor gamma chain: a functional component of the			
RT	interleukin-4 receptor.";			
RL	Science 262:1880-1883(1993).			
RN	[6]			
RP	IDENTIFICATION AS A IL-7R SUBUNIT.			

34	27	79.4	428	1 AROA_CAMJE	P52312 campylobact
35	27	79.4	517	1 MB13_YEAST	O92247 saccharomyc
36	27	79.4	563	1 YCU4_YEAST	P25355 saccharomyc
37	27	79.4	583	1 NIR_BEYVE	P38500 betula verr
38	27	79.4	591	1 ARSF_HUMAN	P54793 homo sapien
39	27	79.4	594	1 NIR_SPIOL	P05314 spinacia ol
40	27	79.4	612	1 Y146_RICPR	O92416 rickettsia
41	27	79.4	716	1 HGFL_MOUSE	P26928 mus musculu
42	27	79.4	726	1 BFL1_HUMAN	P50747 h biotin--p
43	27	79.4	795	1 SYER_BUCAI	P57230 buchera ap
44	27	79.4	878	1 YBM5_SCHPO	O10337 schizosacch
45	27	79.4	890	1 YOUN_ECOLI	P39838 escherichia

RX MEDLINE-94090316; PubMed-8266077.  
 RA Noguchi M., Nakamura Y., Russell S.M., Ziegler S.F., Tsang M., Cao X.,  
 RA Leonard W.J.;  
 RT "Interleukin-2 receptor gamma chain: a functional component of the  
 RT interleukin-7 receptor.";  
 RL Science 262:1877-1880(1993).  
 RN [17]  
 RP 3D-STRUCTURE MODELING OF 57-248.  
 RX MEDLINE-95111955; PubMed-7529123;  
 RA Babrough P., Hedgecock C.J., Richards M.G.;  
 RT "The interleukin-2 and interleukin-4 receptors studied by molecular  
 RT modelling";  
 RL Structure 2:839-851(1994).  
 RN [18]  
 RP VARIANTS XSCID PHE-115; CYS-240 AND IL2-241.  
 RX MEDLINE-94130970; PubMed-8299698;  
 RA Disanto J.P., Dautry-Varsat A., Certain S., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Interleukin-2 (IL-2) receptor gamma chain mutations in X-linked  
 RT severe combined immunodeficiency disease result in the loss of  
 RT high-affinity IL-2 receptor binding.";  
 RL Eur. J. Immunol. 24:475-479(1994).  
 RN [19]  
 RP VARIANT XSCID LYS-68.  
 RX MEDLINE-94375038; PubMed-8088810;  
 RA Markiewicz S., Subtil A., Dautry-Varsat A., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Detection of three nonsense mutations and one missense mutation in  
 RT the interleukin-2 receptor gamma chain gene in SCIDX1 that  
 RT differentially affect the mRNA processing.";  
 RL Genomics 21:291-293(1994).  
 RN [10]  
 RP VARIANT XSCID HIS-162.  
 RX MEDLINE-94300093; PubMed-8027558;  
 RA Ishii N., Asao H., Kimura Y., Takeshita T., Nakamura M., Tsuchiya S.,  
 RA Kono T., Maeda M., Uchiyama T., Sugamura K.;  
 RT "Impairment of ligand binding and growth signaling of mutant IL-2  
 RT receptor gamma-chains in patients with X-linked severe combined  
 RT immunodeficiency.";  
 RL J. Immunol. 153:1310-1317(1994).  
 RN [11]  
 RP VARIANT XSCID ASN-39.  
 RX MEDLINE-95023933; PubMed-7937790;  
 RA Disanto J.P., Rieux-Laucat F., Dautry-Varsat A., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Defective human interleukin 2 receptor gamma chain in an atypical X  
 RT chromosome-linked severe combined immunodeficiency with peripheral T  
 RT cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9466-9470(1994).  
 RN [12]  
 RP VARIANTS XSCID CYS-226 AND HIS-226.  
 RX MEDLINE-95397841; PubMed-7668284;  
 RA Pepper A.E., Buckley R.H., Small T.N., Puck J.M.;  
 RT "Two mutational hotspots in the interleukin-2 receptor gamma chain  
 RT gene causing human X-linked severe combined immunodeficiency.";  
 RL Am. J. Hum. Genet. 57:564-571(1995).  
 RN [13]  
 RP VARIANT XSCID SER-183.  
 RX MEDLINE-96013903; PubMed-7557965;  
 RA Clark P.A., Lester T., Genet S., Jones A.M., Hendriks R.,  
 RA Levinsky R.L., Kinon C.;  
 RT "Screening for mutations causing X-linked severe combined  
 RT immunodeficiency in the IL-2R gamma chain gene by single-strand  
 RT conformation polymorphism analysis.";  
 RL Hum. Genet. 96:427-432(1995).  
 RN [14]  
 RP VARIANT XSCID GLN-237 G-H-M INS.  
 RX MEDLINE-95164726; PubMed-7860773;  
 RA Puck J.M., Pepper A.E., Bedard P.-M., Laframboise R.;  
 RT "Female germ line mosaicism as the origin of a unique IL-2 receptor  
 RT gamma-chain mutation causing X-linked severe combined  
 RT immunodeficiency.";  
 RL J. Clin. Invest. 95:895-899(1995).

RN [15]  
 RP VARIANT XSCID GLN-293.  
 RX MEDLINE-95190013; PubMed-7883965;  
 RA Schmalstieg F.C., Leonard W.J., Noguchi M., Berg M., Rudloff H.E.,  
 RA Denney R.M., Dave S.K., Brooks E.G., Goldman A.S.;  
 RT "Missense mutation in exon 7 of the common gamma chain gene causes a  
 RT moderate form of X-linked combined immunodeficiency.";  
 RL J. Clin. Invest. 95:1169-1173(1995).  
 RN [16]  
 RP VARIANT XSCID ARG-115.  
 RX MEDLINE-97042245; PubMed-8900089;  
 RA Stephan V., Mann V., Le Delst F., Dirksen U., Broeker B.,  
 RA Mueller-Fleckenstein I., Horneff G., Schroten H., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Atypical X-linked severe combined immunodeficiency due to possible  
 RT spontaneous reversion of the genetic defect in T cells.";  
 RL New Engl. J. Med. 335:1563-1567(1996).  
 RN [17]  
 RP VARIANT XSCID GLN-285.  
 RX MEDLINE-97295088; PubMed-9150740;  
 RA Jones A.M., Clark P.A., Katz F., Genet S., McMahon C., Alterman L.,  
 RA Cant A., Kinon C.;  
 RT "B-cell-negative severe combined immunodeficiency associated with a  
 RT common gamma chain mutation.";  
 RL Hum. Genet. 99:677-680(1997).  
 RN [18]  
 RP VARIANT XSCID CYS-222.  
 RX MEDLINE-98064061; PubMed-9399950;  
 RA Sharfe N., Shahar M., Roifman C.M.;  
 RT "An interleukin-2 receptor gamma chain mutation with normal thymus  
 RT morphology.";  
 RL J. Clin. Invest. 100:3036-3043(1997).  
 RN [19]  
 RP FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF  
 CC INTERLEUKINS.  
 CC -1- FUNCTION: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND  
 CC -1- SUBUNIT: THE IL-13 RECEPTOR.  
 CC -1- PROBABLY ALSO THE IL-13 RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A SEVERE COMBINED  
 CC IMMUNODEFICIENCY, WHICH IS KNOWN AS AGAMMAGLOBULINEMIA, SWISS TYPE  
 CC OR X-LINKED SEVERE COMBINED IMMUNODEFICIENCY DISEASE (XSCID).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -1- DATABASE: NAME=PROV: NOTE=CD guide cd132 entry.  
 CC NAME="http://www.ncbi.nlm.nih.gov/prov/cd/cd132.htm".  
 CC -1- DATABASE: NAME=IL2Gdbase; NOTE=X-linked SCID mutation database;  
 CC WWW="http://www.nhgrl.nih.gov/DIR/GMNB/SCID/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-ch](mailto:license@sib-ch)).  
 CC -----  
 DR EMBL: D11086; BAA01857.1; -;  
 DR EMBL: L12183; AAA59145.1; -;  
 DR EMBL: L12178; AAA59145.1; JOINED.  
 DR EMBL: L12176; AAA59145.1; JOINED.  
 DR EMBL: L12177; AAA59145.1; JOINED.  
 DR EMBL: L12179; AAA59145.1; JOINED.  
 DR EMBL: L12180; AAA59145.1; JOINED.  
 DR EMBL: L12181; AAA59145.1; JOINED.  
 DR EMBL: L12182; AAA59145.1; JOINED.  
 DR EMBL: L19546; AAC37524.1; -;  
 DR PIR: A42565; A42565.  
 DR PDB: 1IIM; 26-JAN-95.  
 DR PDB: 1ILN; 26-JAN-95.  
 DR MIM: 308380; -;  
 DR MIM: 300400; -;  
 DR InterPro: IPR002996; CRA.  
 DR InterPro: IPR003961; FN.III.  
 DR InterPro: IPR003531; Hematopo\_receptor\_S\_F1.

Query Match 100.0%; Score 34; DB 1; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 DB 148 LQNLVIP 154

RESULT 2  
 CYRG\_MOUSE STANDARD; PRT; 369 AA.

AC P34902;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Cytokine receptor common gamma chain precursor (Gamma-C)  
 DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).  
 GN IL2RG.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RA MEDLINE-93377575; PubMed-8503926;  
 RA Kunka S., Kondo M., Takeshita T., Asao H., Nakamura M., Sugamura K.;  
 RT "Cloning of the mouse interleukin 2 receptor gamma chain:  
 RT demonstration of functional differences between the mouse and human  
 RT receptors.";  
 RL Biochem. Biophys. Res. Commun. 193:356-363(1993).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CBA/CA.  
 RA MEDLINE-93391374; PubMed-8378320;  
 RA Cao X., Kozak C.A., Liu Y.J., Noguchi M., O'Connell E., Leonard W.J.;  
 RT "Characterization of cDNAs encoding the murine interleukin 2 receptor  
 RT (IL-2R) gamma chain: chromosomal mapping and tissue specificity of  
 RT IL-2R gamma chain expression.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8464-8468(1993).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RA MEDLINE-9336191; PubMed-8359699;  
 RA Kobayashi N., Nakagawa S., Minami Y., Taniguchi T., Kono T.;  
 RT "Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor  
 RT gamma.";  
 RL gene 130:303-304(1993).  
 RN [4]

RP SEQUENCE FROM N.A.  
 RA MEDLINE-95104285; PubMed-7805729;  
 RA Disanto J.P., Certain S., Wilson A., Macdonald H.R., Avner P.,  
 RA Fischer A., de Saint Basile G.;  
 RT "The murine interleukin-2 receptor gamma chain gene: organization,  
 RT chromosomal localization and expression in the adult thymus.";  
 RL Eur. J. Immunol. 24:3014-3018(1994).  
 RN [5]

RP SEQUENCE FROM N.A.  
 RC STRAIN=B6.S;  
 RA MEDLINE-96341745; PubMed-8750189;  
 RA Chiu R.K., Droll A., Cooper D.L., Dougherty S.T., Dirks J.F.,  
 RA Dougherty G.J.;  
 RT "Molecular mechanisms regulating the hyaluronan binding activity of  
 RT the adhesion protein CD44.";  
 RL J. Neurooncol. 26:231-239(1995).

CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF  
 CC INTERLEUKINS.  
 CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND  
 CC PROBABLY ALSO THE IL-13 RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: D13821; BAA02974.1; -  
 CC EMBL: U21795; AAA64279.1; -  
 CC EMBL: D13565; BAA02760.1; -  
 CC EMBL: L20046; AAA39286.1; -  
 CC EMBL: S75852; AAB32904.1; -  
 CC EMBL: S75844; AAB32904.1; JOINED.  
 CC EMBL: S75845; AAB32904.1; JOINED.  
 CC EMBL: S75847; AAB32904.1; JOINED.  
 CC EMBL: S75848; AAB32904.1; JOINED.  
 CC EMBL: S75849; AAB32904.1; JOINED.  
 CC EMBL: S75850; AAB32904.1; JOINED.  
 CC EMBL: S75851; AAB32904.1; JOINED.  
 CC EMBL: X75337; CAA53085.1; -  
 CC PIR: JN0592; JN0592.  
 CC PIR: JN0775; JN0775.  
 CC HSSP: P31785; ILLN.  
 CC MGD: MGI:96551; I12F9.  
 CC InterPro: IPR002996; CRLA.  
 CC InterPro: IPR003961; FN.III.  
 CC InterPro: IPR003531; Hematopo\_receptor\_S\_F1.  
 CC Pfam: PF00041; fn3; 1.  
 CC SMART: SM00060; fn3; 1.  
 CC KEGG: K01355; HEMATOPO\_REC\_S\_F1; 1.  
 CC KW Receptor; Transmembrane; Glycoprotein; Signal.  
 CC FT SIGNAL 1 22  
 CC FT CHAIN 23 369  
 CC FT DOMAIN 23 263  
 CC FT TRANSMEM 264 284  
 CC FT DOMAIN 285 369  
 CC FT DOMAIN 151 250  
 CC FT DISULFID 62 72  
 CC FT DISULFID 102 115  
 CC FT CARBOHYD 71 71  
 CC FT CARBOHYD 75 75  
 CC FT CARBOHYD 84 84  
 CC FT CARBOHYD 96 96  
 CC FT CARBOHYD 159 159  
 CC FT CARBOHYD 164 164  
 CC SQ SEQUENCE 369 AA; 42241 MW; CB2D5AB459077AC7 CRC64;

Query Match 100.0%; Score 34; DB 1; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 DB 148 LQNLVIP 154

RESULT 3  
 CYRG\_CANFA STANDARD; PRT; 373 AA.

AC P40321;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Cytokine receptor common gamma chain precursor (Gamma-C)  
 DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).  
 GN IL2RG.

OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE-Spleen;
RX MEDLINE=95130114; PubMed=7829104;
RA Henthorn P.S., Somberg R.L., Flamlant V.M., Puck J.M., Patterson D.F.,
RA Felsburg P.J.;
RT "IL-2R gamma gene microdeletion demonstrates that canine X-linked
RT severe combined immunodeficiency is a homologue of the human
RT disease."
RL Genomics 23:69-74(1994).
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A CANINE X-LINKED
CC SEVERE COMBINED IMMUNODEFICIENCY.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U04361; AAC48403.1; -
DR HSSP: P31785; IILN.
DR InterPro: IPR002996; CRIA.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003531; Hematopo_receptor_S_F1.
DR Pfam: PF00041; FN3; 1.
DR SMART: SM00060; FN3; 1.
DR PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
KM Receptor: Transmembrane; Glycoprotein; signal.
FT CHAIN 1 373
FT SIGNAL 23 22
FT DOMAIN 23 261
FT TRANSMEM 262 283
FT DOMAIN 284 373
FT DOMAIN 151 249
FT DISULFID 62 72
FT CARBOHYD 102 115
FT CARBOHYD 24 24
FT CARBOHYD 71 71
FT CARBOHYD 75 75
FT CARBOHYD 84 84
FT CARBOHYD 159 159
FT CARBOHYD 164 164
FT CARBOHYD 249 249
SQ SEQUENCE 373 AA; 42516 MW; 03A0DE1F8B089DB8 CRC64;

Query Match 100.0%; Score 34; DB 1; Length 373;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 148 LQNLVIP 154

RESULT 4
GATA_BORBU
ID GATA_BORBU STANDARD; PRT; 496 AA.
AC 051317;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutaryl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
DE subunit A).
GN GATA OR BB0342.

```

```

OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
NCBI_Taxid=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kierulff A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uetrichack T., Watney L., McDonald L., Arlanch P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSMUTATION OF MISACLYLATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE -> ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001140; AAC6715.1; -
DR TIGR: BB0342; -
DR InterPro: IPR000120; Amidase.
DR Pfam: PF01425; Amidase; 1.
DR PROSITE: PS00571; AMIDASES; 1.
KM Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 496 AA; 55544 MW; D375DEE20D42710 CRC64;

Query Match 97.1%; Score 33; DB 1; Length 496;
Best Local Similarity 85.7%; Pred. No. 5.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 395 LQNLVIP 401

RESULT 5
VCAP_HSV11
ID VCAP_HSV11 STANDARD; PRT; 1374 AA.
AC P06491;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major capsid protein (MCP) (capsid protein VP5).
DE UL19.
GN Herpes simplex virus (type 1 / strain 17).
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_Taxid=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of

```

```

RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=87010565; PubMed=3020164;
RT Davison B.A.J., Scott J.E.;
RT "DNA sequence of the major capsid protein gene of herpes simplex
RL virus type 1."
RL J. Gen. Virol. 67:2279-2286(1986).
RN [3]
RP SEQUENCE OF 202-211 AND 607-616.
RA MEDLINE=93019027; PubMed=1328483;
RT Davison M.D., Rixon F.J., Davison A.J.;
RT "Identification of genes encoding two capsid proteins (VP24 and VP26)
RT of herpes simplex virus type 1."
RL J. Gen. Virol. 73:2709-2713(1992).
CC -1- FUNCTION: MAJOR PROTEIN OF THE ICOSAHEDRAL CAPSID.
CC -1- SUBUNIT: MAIN COMPONENT OF THE HEXAVALENT, AND PROBABLY THE
CC PENTAVALENT CAPSOMERES.
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES MAJOR CAPSID PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10879; BAA01665.1; -
DR EMBL: X14112; CAA32332.1; -
DR EMBL: X04467; CAA28154.1; -
DR PIR: A27239; VCBEB7.
DR PIR: A30084; A30084.
DR InterPro: IPR000912; Herpes_MCP.
DR Pfam: PF03122; Herpes_MCP.1.
DR PRINTS: PR00235; HSVCAPSIDMCP.
KM Coat protein.
SO SEQUENCE 1374 AA; 149083 MW; 859C76E2EAD05B7 CRC64;

Query Match
Best Local Similarity 91.2%; Score 31; DB 1; Length 1374;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
Db 839 LQNLVIP 845

RESULT 6
YADB_SCHPO STANDARD; PRT; 267 AA.
AC 009836;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 31.4 kDa protein C4G8.11C in chromosome I.
GN SPAC4G8.11C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO YEAST ATP10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z56276; CAA91212.1; -
KM Hypothetical protein.
SO SEQUENCE 267 AA; 31381 MW; B38390F48D4D835C CRC64;

Query Match
Best Local Similarity 88.2%; Score 30; DB 1; Length 267;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
Db 6 LKNLVIP 12

RESULT 7
VP40_MABVM STANDARD; PRT; 283 AA.
ID VP40_MABVM
AC P35260;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Matrix protein VP40.
GN VP40.
OS Marburg virus (strain Musoke).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Filovirus.
OX NCBI_TaxID=33727;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92327834; PubMed=1626422;
RA Feldmann H., Muehberger E., Randolph A., Will C., Killey M.P.,
RA Sanchez A., Klenk H.D.;
RT "Marburg virus, a filovirus: messenger RNAs, gene order, and
RT regulatory elements of the replication cycle."
RL Virus Res. 24:1-19(1992).
RN [2]
RP REVISIONS.
RA Feldmann H.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z12132; CAA78116.1; ALT_TERM.
KM Matrix protein.
SO SEQUENCE 283 AA; 31653 MW; C338D448B2345265 CRC64;

Query Match
Best Local Similarity 88.2%; Score 30; DB 1; Length 283;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7
Db 146 LQNLVIP 152

RESULT 8
VP40_MABVP STANDARD; PRT; 303 AA.
ID VP40_MABVP
AC 003040;

```

```

DT 01-FEB-1994 (Rel. 28, Last Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Matrix protein VP40 (VP3).
GN VP40.
OS Marburg virus (strain Popp).
OS VIRUSES: ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Filoviridae.
ON NCBI_TaxID=33728;
RX MEDLINE=93245956; PubMed=8482365;
RA Bukreyev A.A., Volchkov V.E., Bilinov V.M., Netesov S.V.;
RT "The VP35 and VP40 proteins of filoviruses. Homology between Marburg
and Ebola viruses."
RL FEBS Lett. 322:41-46(1993).
RN [12]
RP SEQUENCE FROM N.A.
RX MEDLINE=96028047; PubMed=7487490;
RA Bukreyev A.A., Volchkov V.E., Bilinov V.M., Dryga S.A., Netesov S.V.;
RT "The complete nucleotide sequence of the Popp (1967) strain of Marburg
virus: a comparison with the Musoke (1980) strain."
RL Arch. Virol. 140:1589-1600(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64406; CAA45749.1; -
DR EMBL; Z29337; CAA82538.1; -
DR PIR; S32583; S32583.
DR PIR; S28569; S28569.
DR PIR; S44051; S44051.
KM Matrix protein.
SQ SEQUENCE 303 AA; 33734 MW; 4AB3B45402C961DD CRC64;

Query Match      88.2%; Score 30; DB 1; Length 303;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 146 LQNLVIP 152

RESULT 9
ID VAL1_PYMV STANDARD; PRT; 361 AA.
AC P27258;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE AL1 protein.
OS Potato yellow mosaic virus (isolate Venezuela).
OC VIRUSES: ssDNA viruses; Geminiviridae; Begomovirus.
ON NCBI_TaxID=10828;
RX MEDLINE=91311403; PubMed=1856690;
RA Coutts R.H.A., Coffin R.S., Roberts E.J.F., Hamilton W.D.O.;
RT "The nucleotide sequence of the infectious cloned DNA components of
RT potato yellow mosaic virus."
RL J. Gen. Virol. 72:1515-1520(1991).
CC -----
CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00940; BAA00782.1; -
DR PIR; J00364; QOCVPT.
DR InterPro; IPR001191; Gemini1_AL1.
DR Pfam; PF00799; Gemini1_AL1.1.
DR PRINTS; PR00227; GEMCOATL1.
DR Prodom; PD000736; Gemini1_AL1.1.
KW ATP-binding.
FT NP_BIND 222 229 ATP (POTENTIAL).
SQ SEQUENCE 361 AA; 40850 MW; 5627A33BF1264383 CRC64;

Query Match      88.2%; Score 30; DB 1; Length 361;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 32 LQNLVIP 38

RESULT 10
ID YMB8_YEAST STANDARD; PRT; 553 AA.
AC 004847;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 64.4 kDa protein in PER11-TIF11 intergenic region.
GN YMR258C OR YM9920.12C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
ON NCBI_TaxID=4932;
RX MEDLINE=248639; CAA88585.1; -
DR SGD; S0004871; YMR258C.
KM Hypothetical protein.
SQ SEQUENCE 553 AA; 64405 MW; 2F764964A6C23EC2 CRC64;

Query Match      88.2%; Score 30; DB 1; Length 553;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQNLVIP 7
Db 192 LQNLVIP 198

RESULT 11
ID CYRG_BOVIN STANDARD; PRT; 379 AA.
AC 095118;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Cytokine receptor common gamma chain precursor (Gamma-C)  
 DE (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).  
 GN IL2RG.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 ON NCBI\_TaxID=9913;  
 RX MEDLINE-96268473; PubMed-8672241;  
 RA Yoo J., Stone R.T., Solinas-Toldo S., Fries R., Beattie C.W.;  
 RT "Cloning and chromosomal mapping of bovine interleukin-2 receptor  
 gamma gene."  
 CC DNA Cell Biol. 15:453-459(1996).  
 CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF  
 CC INTERLEUKINS.  
 CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND  
 CC PROBABLY ALSO THE IL-13 RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U33748; AAB07812.1; -  
 CC HSSP: P31785; IILN.  
 DR InterPro: IPR002996; CRIA.  
 DR InterPro: IPR003961; FN.III.  
 DR InterPro: IPR003531; Hematopo\_receptor\_S\_F1.  
 DR Pfam: PF00041; FN3; 1.  
 DR SMART: SM00060; FN3; 1.  
 DR PROSITE: PS01355; HEMATOPO\_REC\_S\_F1; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 1 379  
 FT DOMAIN 23 379 CYTOKINE RECEPTOR COMMON GAMMA CHAIN.  
 FT TRANSMEM 270 290 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 291 379 POTENTIAL.  
 FT DOMAIN 158 256 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 68 78 FIBRONECTIN TYPE-III.  
 FT DISULFID 109 122 POTENTIAL.  
 FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 379 AA; 43037 MW; 33CFAD9C9B032178 CRC64;

Query Match 85.3%; Score 29; DB 1; Length 379;  
 Best Local Similarity 85.7%; Pred. No. 35;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 Db 155 LQDLVIP 161

RESULT 12  
 YLEU\_DEBOC  
 AC P48012; STANDARD; PRT; 379 AA.  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Hypothetical 43.6 kDa protein.  
 OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 ON NCBI\_TaxID=27300;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 26077 / R91;  
 RX MEDLINE-95321019; PubMed-7597851;  
 RA Isernante D., Verachert H.;  
 RT "Cloning and sequencing of the LEU2 homologue gene of Schwanniomyces  
 RT occidentalis."  
 RL Yeast 11:467-473(1995).  
 CC -1- SIMILARITY: TO YEAST BUD3.  
 CC -1- CAUTION: IS INDICATED BY REF.1 TO BE A 3-ISOPROPYLMALATE  
 CC DEHYDROGENASE (EC 1.1.1.85) (LEU2), BUT THIS IS MOST PROBABLY A  
 CC MISASSIGNMENT.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X79823; CAA56224.1; -  
 CC InterPro: IPR000219; RhogEF.  
 DR SMART: SM00325; RhogEF; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 379 AA; 43586 MW; 6C846CD0B53B0BC CRC64;

Query Match 85.3%; Score 29; DB 1; Length 379;  
 Best Local Similarity 71.4%; Pred. No. 35;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LQNLVIP 7  
 Db 46 LQNLVIP 52

RESULT 13  
 YMA2\_CAEEL  
 ID YMA2\_CAEEL STANDARD; PRT; 535 AA.  
 AC P34447;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 56.5 kDa protein F54F2.2 in chromosome III.  
 GN F54F2.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 ON NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkhen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Spoat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL Nature 368:32-38(1994).

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L23645; AAA28046.1; -
DR PIR: S44827; S44827.
DR Morphpep: F54F2.2; CE00195.
DR Hypothetical protein.
DR KW SEQUENCE 535 AA; 56492 MW; 731E133D327AD9A CRC64;
SQ
Query Match      85.3%; Score 29; DB 1; Length 535;
Best Local Similarity 85.7%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LQNLVIP 7
Db 150 LQNLVIP 156
RESULT 14
C2AC_BACTU STANDARD; PRT; 622 AA.
AC 045743;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry2Ac (Insecticidal delta-endotoxin
DE CryIIA(C)) (Crystalline entomocidal protoxin) (70 kDa crystal protein).
GN CRY2AC OR CRYIIA(C) OR CRYIIC.
OS Bacillus thuringiensis.
OC plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SHANGHAI 1 / S-1;
RX MEDLINE-91340086; PubMed-1651878;
RA Wu D., Cao X.L., Bai Y.Y., Aronson A.I.;
RT "Sequence of an operon containing a novel delta-endotoxin gene from
RT Bacillus thuringiensis."
RL FEMS Microbiol. Lett. 65:31-36(1991).
CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF LEPIDOPTERAN LARVAE. HAS LOW ACTIVITY ON
CC DIPTERAN LARVAE.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57252; CAA40536.1; -
DR InterPro: IPR001178; Endotoxin.
DR Pfam: PF00555; endotoxin.1.
DR Toxin; Sporulation; Plasmid.
DR KW SEQUENCE 622 AA; 69729 MW; E0B5FAD37BF8299 CRC64;
SQ

```

```

Query Match      85.3%; Score 29; DB 1; Length 622;
Best Local Similarity 71.4%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LQNLVIP 7
Db 77 LQNLVIP 83
RESULT 15
ALIB_STRPN STANDARD; PRT; 652 AA.
ID ALIB_STRPN
AC Q51933;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oligopeptide-binding protein alib precursor.
GN ALIB OR SP1527.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R800;
RX MEDLINE-94328326; PubMed-8051706;
RA Allouin G., de Philip P., Claverys J.-P.;
RT "Three highly homologous membrane-bound lipoproteins participate in
RT oligopeptide transport by the Aml system of the Gram-positive
RT Streptococcus pneumoniae."
RL J. Mol. Biol. 241:44-58(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE-21357209; PubMed-11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfe E., Khouli H., Wolf A.M., Ueberback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angibol S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR OLIGOPEPTIDES. PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z16082; CAA78896.1; -
DR EMBL: AE007448; AA75616.1; -
DR TIGR: SP1527;
DR InterPro: IPR000914; SBP_bac-5.
DR Pfam: PF00496; SBP_bac-5; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
DR KW Peptide transport; Transport; Membrane; Lipoprotein; Signal;
DR Complete proteome.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 652 OLIGOPEPTIDE-BINDING PROTEIN ALIB.
FT LIPID 25 25 N-ACYL DIGLYCERIDE (PROBABLE).

```



FT CONFLICT 55 55 A -> R (IN REF. 1).  
 FT CONFLICT 79 80 SL -> HI (IN REF. 1).  
 FT CONFLICT 123 124 LO -> FE (IN REF. 1).  
 FT CONFLICT 501 501 G -> E (IN REF. 1).  
 SQ SEQUENCE 652 AA; 72562 MM; 169B67FD78CF0CF CRC64;

Query Match 85.3%; Score 29; DB 1; Length 652;  
 Best Local Similarity 71.4%; Pred. No. 61;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVIP 7  
 1:111:1  
 DB 363 LRNLVVP 369

Search completed: August 20, 2002, 11:33:02  
 Job time: 1451 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:55 ; Search time 191.14 Seconds  
(without alignments)  
6.335 Million cell updates/sec

Title: US-09-824-286-15  
Perfect score: 34  
Sequence: 1 LONLVIP 7

Scoring table: BLASTSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp-organellar:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp\_bacteriap:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	273	2	093DQ0	093dQ0 uncultured
2	34	100.0	273	2	093DP7	093dP7 uncultured
3	32	94.1	380	5	090943	09u943 locusta mlg
4	31	91.2	358	13	09W616	09w616 brachydanio
5	31	91.2	358	13	09PW55	09pw55 brachydanio
6	31	91.2	1374	12	P89442	P89442 herpes simp
7	30	88.2	225	12	09WHF8	09whf8 sinaloa tom
8	30	88.2	235	12	039178	039178 geminivirid
9	30	88.2	346	8	021187	021187 lamna nasus
10	30	88.2	362	13	09W617	09w617 brachydanio
11	30	88.2	372	12	09PYV2	09pyv2 xestia c-ni
12	30	88.2	373	11	09CXB2	09cxb2 mus musculus
13	30	88.2	634	5	044116	044116 drosophila
14	30	88.2	634	5	09V978	09v978 drosophila
15	30	88.2	689	5	018668	018668 caenorhabdi
16	30	88.2	1007	5	096848	096848 dictyosteli

17	30	88.2	1007	5	Q23897	Q23897 dictyosteli
18	29	85.3	171	5	094184	094184 caenorhabdi
19	29	85.3	184	10	09XFE2	09xfe2 arabidopsis
20	29	85.3	217	10	040292	040292 beta vulgar
21	29	85.3	246	16	0990W7	0990w7 staphylococ
22	29	85.3	261	10	09SMW1	09smw1 arabidopsis
23	29	85.3	345	10	09SK24	09sk24 arabidopsis
24	29	85.3	346	8	021195	021195 megachasma
25	29	85.3	346	8	021676	021676 alopias pel
26	29	85.3	346	8	021129	021129 alopias vul
27	29	85.3	461	10	09SDP4	09sdp4 allium cepa
28	29	85.3	513	5	09XW90	09xw90 caenorhabdi
29	29	85.3	537	10	09S2V0	09s2v0 arabidopsis
30	29	85.3	633	4	096CT8	096ct8 homo sapien
31	29	85.3	641	4	094927	094927 homo sapien
32	29	85.3	661	17	052001	052001 halobacteri
33	29	85.3	662	16	09ZDU7	09zdu7 rickettsia
34	29	85.3	662	16	0921W1	0921w1 rickettsia
35	29	85.3	744	16	09PG19	09pg19 xylella fas
36	29	85.3	839	5	010652	010652 caenorhabdi
37	29	85.3	980	5	09HGI1	09hgi1 halocynthia
38	28	82.4	79	2	099223	099223 lactococcus
39	28	82.4	99	15	0999N3	0999n3 human immun
40	28	82.4	151	17	0972F9	0972f9 sulfolobus
41	28	82.4	174	16	09KGG7	09k97 bacillus ha
42	28	82.4	188	10	09ZTD0	09ztd0 arabidopsis
43	28	82.4	216	2	P95356	P95356 neisseria g
44	28	82.4	216	16	09KIR3	09kir3 neisseria m
45	28	82.4	216	16	09JW11	09jw11 neisseria m

## ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	273 AA.
093DQ0	093DQ0			
AC	093DQ0:			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DE	2,5-DIKETO-D-GLUCONIC ACID REDUCTASE.			
OS	uncultured bacterium.			
OC	Bacteria; environmental samples.			
OX	NCBI_TaxID=77133;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MDLINE=21417047; PubMed=11526025;			
RA	Eschenfeldt W.H., Stols L., Rosenbaum H., Khambatla Z.S.,			
RA	Quake-Randall E., Wu S., Kilgore D.C., Trent J.D., Donnelly M.I.,			
RT	"Dna from uncultured organisms as a source of 2,5-diketo-D-gluconic			
RT	acid reductases."			
RL	Appl. Environ. Microbiol. 67:4206-4214(2001).			
DR	EMBL: AF385141; AAK70425.1; -			
SO	SEQUENCE 273 AA; 30864 MW; B4635B614DB9F3B3 CRC64;			
Query Match 100.0%; Score 34; DB 2; Length 273;				
Best Local Similarity 100.0%; Pred. No. 8.7;				
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 LONLVIP 7			
DB	223 LONLVIP 229			
RESULT	2			
ID	093DP7	PRELIMINARY;	PRT;	273 AA.
AC	093DP7:			
DT	01-DEC-2001 (Tremblrel. 19, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			

DE 2,5-DIKETO-D-GLUCONIC ACID REDUCTASE.  
 OS uncultured bacterium.  
 OC Bacteria; environmental samples.  
 OX NCBI\_Taxid=77133;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21417047; PubMed=11526025;  
 RX Eschenfeldt W.H., Stols L., Rosenbaum H., Khambatta Z.S.,  
 RA Qualte-Randall E., Wu S., Kilgore D.C., Trent J.D., Donnelly M.I.;  
 RT "Dna from uncultured organisms as a source of 2,5-diketo-D-gluconic  
 RT acid reductases.";  
 RL Appl. Environ. Microbiol. 67:4206-4214(2001).  
 DR EMBL: AF385143; AK70428.1; .  
 SQ SEQUENCE 273 AA; 30879 MW; 8E73501027A93363 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 273;  
 Best Local Similarity 100.0%; Pred. No. 8.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LQNLVIP 7  
 |||||  
 Db 223 LQNLVIP 229

RESULT 3  
 ID Q9U943 PRELIMINARY; PRT; 3380 AA.  
 AC Q9U943;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE APOLIPOPHORIN PROTEIN PRECURSOR.  
 OS Locusta migratoria (Migratory locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Orthoptera; Orthoptera; Caelifera;  
 OC Acridomorpha; Acridoidea; Acrididae; Locusta.  
 OX NCBI\_Taxid=7004;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FAI BODY;  
 RA Bogerd J., Bablin P.J., Koolman F.P., Van der Horst D.J.;  
 RT "Cloning and expression of cDNA encoding the apolipoprotein precursor  
 RT protein mRNA of Locusta migratoria.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ130944; CAB51918.1; .  
 DR InterPro: IPR003662; sub\_transprot.  
 DR InterPro: IPR001747; Vitellogenin\_N.  
 DR InterPro: IPR001846; Vwd.  
 DR Pfam: PF01347; Vitellogenin\_N; 3.  
 DR Pfam: PF00094; Vwd; 1.  
 DR SMART: SM00216; VWD; 1.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
 KW Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 716 APOLIPOPHORIN II.  
 FT CHAIN 721 3380 APOLIPOPHORIN I.  
 SQ SEQUENCE 3380 AA; 371719 MW; DF3EE9B46EB8FDB5 CRC64;

Query Match 94.1%; Score 32; DB 5; Length 3380;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LQNLVIP 7  
 |||||  
 Db 2670 LQNLVIP 2676

RESULT 4  
 ID Q9W616 PRELIMINARY; PRT; 358 AA.  
 AC Q9W616;

DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE SIGNALING MOLECULE LEFTY1.  
 GN LEFTY.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Danio.  
 OX NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99307087; PubMed=10375514;  
 RA Bisgrove B.W., Essner J.J., Yost H.J.;  
 RT "Regulation of midline development by antagonism of lefty and nodal  
 RT signaling.";  
 RL Development 126:3253-3262(1999).  
 CC -I- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.  
 DR EMBL: AF132444; AAD34388.1; .  
 DR HSSP: P01137; IKLA.  
 DR ZFIN: ZDB-GENE-990630-10; Ifcl.  
 DR InterPro: IPR001839; TGF-beta.  
 DR InterPro: IPR001111; TGFb\_N.  
 DR Pfam: PF00019; TGF-beta; 1.  
 DR Pfam: PF00688; TGFb\_propeptide; 1.  
 DR ProDom: PD000357; TGF-beta; 1.  
 DR SMART: SM00204; TGFb; 1.  
 DR PROSITE: PS00250; TGF-BETA; 1.  
 KW Glycoprotein.  
 SQ SEQUENCE 358 AA; 41018 MW; F7A86FE828092A8 CRC64;

Query Match 91.2%; Score 31; DB 13; Length 358;  
 Best Local Similarity 85.7%; Pred. No. 56;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LQNLVIP 7  
 |||||  
 Db 47 LQNLVIP 53

RESULT 5  
 ID Q9PW55 PRELIMINARY; PRT; 358 AA.  
 AC Q9PW55;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE ANTIYIN.  
 GN LEFTY.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Danio.  
 OX NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99065494; PubMed=9847237;  
 RA Thisse C., Thisse B.;  
 RT "Antiyn, a novel and divergent member of the Tgfbeta superfamily,  
 RT negatively regulates mesoderm induction.";  
 RL Development 126:229-240(1999).  
 CC -I- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.  
 DR EMBL: AF172089; AAD47114.1; .  
 DR HSSP: P01137; IKLA.  
 DR ZFIN: ZDB-GENE-990630-10; Ifcl.  
 DR InterPro: IPR001839; TGF-beta.  
 DR InterPro: IPR001111; TGFb\_N.  
 DR Pfam: PF00019; TGF-beta; 1.  
 DR Pfam: PF00688; TGFb\_propeptide; 1.  
 DR ProDom: PD000357; TGF-beta; 1.  
 DR SMART: SM00204; TGFb; 1.  
 DR PROSITE: PS00250; TGF-BETA; 1.

KW Glycoprotein.  
SQ SEQUENCE 358 AA; 41041 MW; 74D3036F18D2458F CRC64;

Query Match 91.2%; Score 31; DB 13; Length 358;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVTP 7  
|:|||||  
DB 47 LQNLVTP 53

```

RESULT 6
P89442 PRELIMINARY; PRT; 1374 AA.
AC P89442;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MAJOR CAPSID PROTEIN.
GN UL19.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the short
RT unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons."
RT J. Gen. Virol. 68:19-38(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=90278430; PubMed=2161906;
RA Everett R., Fenwick M.;
RT "Comparative DNA sequence analysis of the host shutoff genes of
RT different strains of herpes simplex virus: type 2 strain HG52 encodes
RT a truncated UL41 product."
RT J. Gen. Virol. 71:1387-1390(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2."
RT J. Gen. Virol. 72:3057-3073(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92356101; PubMed=1322965;
RA Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;
RT "A novel herpes simplex virus gene (UL49A) encodes a putative membrane
RT protein with counterparts in other herpesviruses."
RT J. Gen. Virol. 73:2167-2171(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RA Dolan A.;
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z86099; CAB06743.1;
DR InterPro: IPR000912; Herpes_MCP.
DR Pfam: PF03122; Herpes_MCP.1.
DR PRINTS: PR00235; HSVCAPSIDMCP.
SQ SEQUENCE 1374 AA; 149237 MW; EC8E85351E4C1939 CRC64;
```

Query Match 91.2%; Score 31; DB 12; Length 1374;  
Best Local Similarity 71.4%; Pred. No. 2e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LQNLVTP 7  
|:|||||  
DB 839 LQNLVTP 845

```

RESULT 7
Q9WHF8 PRELIMINARY; PRT; 225 AA.
AC Q9WHF8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE REPLICATION-ASSOCIATED PROTEIN (FRAGMENT).
GN REP.
OS Sinusoa tomato leaf curl geminivirus.
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=71186;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CR2;
RA Karakashian J.P., Nakha M.K., Maxwell D.P., Ramirez P.;
RT "Molecular characterization of tomato-infecting geminiviruses in Costa
RT Rica."
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF131213; AAD34731.1;
DR InterPro: IPR001191; Geminl_AL1.
DR Pfam: PF00799; Geminl_AL1.1.
DR PRINTS: PR00227; GEMCOATAL1.
DR PRODOM: PD000736; Geminl_AL1.1.
FT NON_TER 225
SQ SEQUENCE 225 AA; 25786 MW; 543C268BD052BDC8 CRC64;
```

```

Query Match 88.2%; Score 30; DB 12; Length 225;
Best Local Similarity 85.7%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LQNLVTP 7
|:|||||
DB 32 LQNLVTP 38

RESULT 8
O39178 PRELIMINARY; PRT; 235 AA.
AC O39178;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AC1 (FRAGMENT).
OS Geminiviridae.
OC Viruses; ssDNA viruses.
OX NCBI_TaxID=10811;
RN [1]
RP SEQUENCE FROM N.A.
RA Guzman P., Arredondo C.R., Emmatly D., Portillo R.J., Gilbertson R.L.;
RT "Partial Characterization of Two Whitefly-Transmitted Geminiviruses
RT Infecting Tomatoes in Venezuela."
RT Plant Dis. 81:312-312(1997).
DR EMBL: AF026464; AAB82598.1;
DR InterPro: IPR001191; Geminl_AL1.
DR Pfam: PF00799; Geminl_AL1.1.
DR PRINTS: PR00227; GEMCOATAL1.
DR PRODOM: PD000736; Geminl_AL1.1.
FT NON_TER 235
SQ SEQUENCE 235 AA; 26522 MW; 317E4D370F16D2DB CRC64;
```

```
Query Match      88.2%; Score 30; DB 12; Length 235;
Best Local Similarity 85.7%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 QNLVLP 7
      |||||
Db      32 QNLVLP 38

RESULT 9
ID 021187 PRELIMINARY; PRT; 346 AA.
AC 021187;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 2 (FRAGMENT).
GN NADH2.
OS Lama nasus.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Galeoidea; Lamniformes; Alopiidae;
OC Lama.
OX NCBI_TaxID=7849;
RN [1]
RP SEQUENCE FROM N.A.
RA Naylor G.J.P., Martin A.P., Mattison E.G., Brown W.M.;
RT "Interrelationships of Lamniform Sharks: Testing Phylogenetic
Hypotheses with Sequence Data."
RL (In) Kocher T.D., Stepien C.A. (eds.);
MOLECULAR SYSTEMATICS OF FISHES, pp.195-214, Academic Press,
San Diego, CA, USA (1997).
CC -1- CATALYTIC ACTIVITY: NADH + UBILQUINONE = NAD(+) + UBILQUINOL.
DR EMBL; U91427; AAB63136.1; -.
DR InterPro; IPR001750; Oxidored_q1.
PFam; PF00361; oxidored_q1; 1.
KM Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
FT NON_TER
SQ SEQUENCE 346 AA; 38035 MW; F0A152FF19926F6 CRC64;

Query Match      88.2%; Score 30; DB 8; Length 346;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 QNLVLP 7
      |||||
Db      273 QNLVLP 278

RESULT 10
ID 09W617 PRELIMINARY; PRT; 362 AA.
AC 09W617;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SIGNALING MOLECULE LEFT2.
GN Brachydanio relio (zebrafish) (zebra danio).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99307087; Pubmed=10375514;
RA Blagrove B.W., Essner J.J., Yost H.J.;
RT "Regulation of midline development by antagonism of lefty and nodal
signaling."
RL Development 126:3253-3262(1999).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
DR EMBL; AF132445; AAD34389.1; -.

+ 37
```

```
DR HSP; P08112; 2TGT.
DR ZFIN; ZDB-GENE-990630-11; lft2.
DR InterPro; IPR001839; TGF-beta.
DR InterPro; IPR001111; TGFb_N.
DR Pfam; PF000019; TGF-beta; 1.
DR Pfam; PF006688; TGFb_propeptide; 1.
DR ProDom; PD0000357; TGF-beta; 1.
DR SMART; SM00204; TGFb; 1.
DR PROSITE; PS00250; TGF-BETA; 1.
KW Glycoprotein.
SQ SEQUENCE 362 AA; 41459 MW; 0C12BCB3CEB6007 CRC64;

Query Match      88.2%; Score 30; DB 13; Length 362;
Best Local Similarity 71.4%; Pred. No. 96;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY      1 QNLVLP 7
      |||||
Db      46 QNLVLP 52

RESULT 11
ID 09PYV2 PRELIMINARY; PRT; 372 AA.
AC 09PYV2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ORF91.
OS Xestia c-nigrum granulosis virus (Xncv) (Xestia c-nigrum
granulovirus).
OS Virusess; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX NCBI_TaxID=51677;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99434230; Pubmed=10502508;
RA Hayakawa T., Ko R., Okano K., Seong S.I., Goto C., Maeda S.;
RT "Sequence analysis of the Xestia c-nigrum granulovirus genome."
RL Virology 262:277-297(1999).
DR EMBL; AF162221; AAF05205.1; -.
SQ SEQUENCE 372 AA; 43839 MW; ECE90B7BB79FA772 CRC64;

Query Match      88.2%; Score 30; DB 12; Length 372;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 QNLVLP 7
      |||||
Db      338 QNLVLP 343

RESULT 12
ID 09CXB2 PRELIMINARY; PRT; 373 AA.
AC 09CXB2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 8430417G17RK PROTEIN.
GN 8430417G17RK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LUNG;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
```

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kato K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuhl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,  
 RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK018423; BAB31203.1; -  
 DR MGD: MG1:192300; 8430417G1RLK.  
 DR InterPro: IPR002657; SBF.  
 DR Pfam: PF01758; SBF; 1.  
 SO SEQUENCE 373 AA; 40681 MW; 0902D18506A8AC55 CRC64;

Query Match 88.2%; Score 30; DB 11; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ONLVIP 7  
 - - - - -  
 Db 156 ONLVIP 161

RESULT 13  
 044116 PRELIMINARY; PRT; 634 AA.  
 AC 044116;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VIMAR.  
 GN VIMAR OR CG3572  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lo P.C.H., Frasch M.,  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF034421; AAB87984.1; -  
 DR FlyBase: FBgn0022960; Vimar.  
 DR InterPro: IPR000225; Armadillo.  
 DR Pfam: PF00514; Armadillo\_seg; 5.  
 DR SMART: SM00185; ARM; 3.  
 SO SEQUENCE 634 AA; 70237 MW; 0968BD62DF1455FD CRC64;

Query Match 88.2%; Score 30; DB 5; Length 634;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ONLVIP 7  
 - - - - -  
 Db 385 ONLVIP 391

RESULT 14  
 09V978 PRELIMINARY; PRT; 634 AA.  
 ID 09V978;  
 AC 09V978;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE VIMAR PROTEIN.  
 GN VIMAR OR CG3572.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Beeson K.Y., Bess A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Bertone P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaisai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson K.A., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003790; AAP57418.1; -  
 DR FlyBase: FBgn0022960; Vimar.  
 DR InterPro: IPR000225; Armadillo.  
 DR Pfam: PF00514; Armadillo\_seg; 5.  
 DR SMART: SM00185; ARM; 3.  
 SO SEQUENCE 634 AA; 70253 MW; 337A23DB99A5486B CRC64;

Query Match 88.2%; Score 30; DB 5; Length 634;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ONLVIP 7  
 - - - - -  
 Db 385 ONLVIP 391

RESULT 15  
 018668 PRELIMINARY; PRT; 689 AA.  
 ID 018668;  
 AC 018668;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
DE CA7D12.2 PROTEIN.  
GN CA7D12.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;  
OC Rhabdilitae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA GaJadsty S.;  
RU Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z69902; CAA93767.1; -;  
SQ SEQUENCE 689 AA; 78202 MW; 8CF57ED089C0D49F CRC64;

Query Match 88.2%; Score 30; DB 5; Length 689;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LONLVP 7  
I:||||:  
Db 392 LONLVP 398

Search completed: August 20, 2002, 11:31:56  
Job time: 1470 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:17 ; Search time 187.18 Seconds  
(without alignments)  
2.967 Million cell updates/sec

Title: US-09-824-286-16

Perfect score: 34

Sequence: 1 HCLEH 5

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802: \*  
1: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: \*  
2: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: \*  
3: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: \*  
4: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: \*  
5: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: \*  
6: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: \*  
7: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: \*  
8: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: \*  
9: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: \*  
10: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: \*  
11: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: \*  
12: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: \*  
13: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: \*  
14: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: \*  
15: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: \*  
16: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: \*  
17: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: \*  
18: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: \*  
19: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: \*  
20: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: \*  
21: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: \*  
22: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	5	AAW31653	Human cytokine rec
2	34	100.0	105	AAW31653	Human cytokine rec
3	34	100.0	174	AAW38627	Streptococcus pneu
4	34	100.0	192	AAW11279	Streptococcus pneu
5	34	100.0	229	AAW35440	Pinus radiata cell
6	34	100.0	230	AAW47151	IL-2 receptor gamm
7	34	100.0	230	AAW82934	Interleukin 4 comp
8	34	100.0	252	AAW47150	IL-2 receptor gamm
9	34	100.0	270	AAW54733	Arabidopsis thalia
10	34	100.0	297	AAW81769	Streptococcus pneu
11	34	100.0	347	AAW47149	IL-2 receptor gamm

12	34	100.0	363	22	AAW71682	Bos taurus interle
13	34	100.0	369	15	AAW47148	IL-2 receptor gamm
14	34	100.0	482	19	AAW31646	Human cytokine rec
15	34	100.0	691	21	AAW92202	Fusion polypeptide
16	34	100.0	694	21	AAW92201	Fusion polypeptide
17	34	100.0	694	21	AAW92203	Fusion polypeptide
18	34	100.0	2645	22	AAW20077	Novel human diago
19	32	94.1	105	22	AAW47099	Human colon cancer
20	32	94.1	364	22	AAW60565	Drosophila melanog
21	31	91.2	30	22	AAW79026	Human protein SEQ
22	31	91.2	56	22	AAW80010	Human protein SEQ
23	31	91.2	89	21	AAW33478	Arabidopsis thalia
24	31	91.2	141	21	AAW33477	Arabidopsis thalia
25	31	91.2	148	22	AAW53000	Propionibacterium
26	31	91.2	178	21	AAW18383	Arabidopsis thalia
27	31	91.2	199	11	AAW06498	GST-1 clone encode
28	31	91.2	270	21	AAW18382	Arabidopsis thalia
29	31	91.2	270	21	AAW49119	Arabidopsis thalia
30	31	91.2	317	21	AAW50357	Arabidopsis thalia
31	31	91.2	326	21	AAW50356	Arabidopsis thalia
32	31	91.2	328	21	AAW50355	Arabidopsis thalia
33	31	91.2	330	21	AAW49118	Arabidopsis thalia
34	31	91.2	338	21	AAW18381	Arabidopsis thalia
35	31	91.2	338	21	AAW49117	Arabidopsis thalia
36	31	91.2	425	14	AAW45151	Sequence of mouse
37	31	91.2	425	17	AAW06798	Murine p154. Mus
38	31	91.2	515	25	AAW42522	Arabidopsis thalia
39	31	91.2	748	22	AAW07878	Polypeptide sequen
40	31	91.2	1316	22	AAW39108	Human polypeptide
41	31	91.2	1401	22	AAW40894	Human polypeptide
42	31	91.2	1612	22	AAW71504	Drosophila melanog
43	31	91.2	75	22	AAW24193	Human EST encoded
44	30	88.2	84	22	AAW88953	Human immune/haema
45	30	88.2	141	22	AAW17027	Novel human diagno

#### ALIGNMENTS

RESULT 1	
AAW31653	
ID AAW31653 standard; Peptide: 5 AA.	
XX	
AC AAW31653;	
XX	
DT 21-MAY-1998 (first entry)	
XX	
DE Human cytokine receptor gc chain epitope.	
XX	
KW Cytokine receptor; gamma common chain; gc chain; human;	
KW blocking agent; monoclonal antibody; Cp B8; immunological disease;	
KW myasthenia gravis; Rheumatoid arthritis; lupus; multiple sclerosis;	
KW insulin-dependent diabetes; inflammatory bowel disease;	
KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;	
KW graft versus host disease; psoriasis; immunosuppressive; therapy;	
KW epitope.	
XX	
OS Homo sapiens.	
XX	
PN WO9743416-A1.	
XX	
PD 20-NOV-1997.	
XX	
PF 09-MAY-1997; 97WO-US07870.	
XX	
PR 10-MAY-1996; 96US-0017466.	
XX	
PA (BIOJ ) BIOGEN INC.	
XX	
PI Benjamin CD, Burkiy LC, Hession C, Whitty A;	
XX	
DR WPI: 1998-008885/01.	
XX	

PT Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
XX  
PS Claim 24; Page 85; 11pp; English.  
XX  
XX This peptide comprises an epitope of the human cytokine receptor  
CC common gamma (gc) chain (see AAW31646) that is recognised by  
CC gc blocking agents of the invention. 5 Such epitopes (see  
CC AAW31650-54) have been identified. The invention provides  
CC compositions and methods for inhibiting cytokine signalling using  
CC gc chain blocking agents for the treatment of immunological  
CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,  
CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel  
CC disease, sympathetic ophthalmia, uveitis, allergy, asthma,  
CC parasitic infection, graft vs. host disease or psoriasis. A  
CC preferred gc blocking agent is Mab CP.B8 or its Fab fragment (see  
CC also AAW31647-48).  
XX  
SQ Sequence 5 AA;

Query Match 100.0%; Score 34; DB 19; Length 5;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
| | | | |  
DB 1 hcleh 5

## RESULT 2

AAG77373  
ID AAG77373 standard; Protein; 105 AA.

AC AAG77373;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SPQ ID NO:8139.

DE Human colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma.

KW Homo sapiens.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PE 28-SEP-2000; 2000WO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI: 2001-223537/24.

DR N-PSDB: AAW36780.

PS Claim 11; Page 9461-9462; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patient's own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated P,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAW77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 105 AA;

Query Match 100.0%; Score 34; DB 22; Length 105;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
| | | | |  
DB 9 hcleh 13

## RESULT 3

AAW38627  
ID AAW38627 standard; Protein; 174 AA.

AC AAW38627;

DT 09-NOV-1998 (first entry)

DE Streptococcus pneumoniae protein of unknown function.

KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
KW immunological response; inoculation; antibody production; inhibitor;  
KW T cell immune response; antimicrobial compound; bacterial adhesion;  
KW extracellular matrix protein; protein-mediated cell invasion; wound;  
KW pathogenesis.

KW Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

PN WO9743303-A1.

PD 20-NOV-1997.

PE 14-MAY-1997; 97WO-US07950.

PR 14-MAY-1996; 96US-0017670.

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

PI Stodola RK;

DR WPI: 1998-008793/01.

DR N-PSDB: AAT98673.

PS Novel Streptococcus pneumoniae proteins and related DNA - useful for  
PT diagnosing anti-microbial agents for treatment of bacterial  
PT infections

PS Claim 12; Pages 387-388; 483pp; English.

This sequence represents a Streptococcus pneumoniae protein of  
CC unknown function, and is encoded by a DNA sequence of the invention.  
CC The DNA sequences were isolated from Streptococcus pneumoniae strain  
CC 0100993 (NCIMB 40794). The streptococcus pneumoniae proteins of the  
CC invention can be used to identify compounds which interact with and  
CC inhibit or activate the activity of the proteins. Antagonists can be

CC used to treat diseases caused by *S. pneumoniae* proteins, through genetic  
CC immunisation. They can also be used to induce an immunological response  
CC in a mammal by inoculation with the *S. pneumoniae* proteins or delivery  
CC of the encoding nucleic acids in a vector adequate to produce antibody  
CC and/or T cell immune responses to protect the animal from disease. The  
CC proteins can also be used to identify antimicrobial compounds which are  
CC capable of inhibiting their bioactivity. In particular the proteins of  
CC the invention can be used to prevent adhesion of bacteria to mammalian  
CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
CC block protein-mediated mammalian cell invasion, and to block the normal  
CC progression of pathogenesis in infections initiated other than by the  
CC implantation of in-dwelling devices or other surgical techniques.

CC  
XX  
SQ Sequence 174 AA;

Query Match 100.0%; Score 34; DB 19; Length 174;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
Db 91 hcleh 95

RESULT 4

ID AAY11279 standard; Protein; 192 AA.

XX AAY11279;

DT 20-MAY-1999 (first entry)

XX Streptococcus pneumoniae protein sequence ID NO:389.

DE Streptococcus pneumoniae strain 0100993; vaccine; immune response;

KM Streptococcus pneumoniae infection; pneumococcal.

XX Streptococcus pneumoniae.

PN W09737026-A1.

PD 09-OCT-1997.

PF 01-APR-1997; 97WO-US05306.

PR 22-AUG-1996; 96US-0025788.

PR 02-APR-1996; 96US-0014690.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PI Black MT, Hodgson JE, Knowles DUC, Nicholas RO;

PI Stodola RK;

DR WPI; 1997-503111/46.

DR N-PDB; AAX30862.

PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in

PT vaccines, drug screening, etc

PS Claim 6; Page 294-295; 354pp; English.

CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from

CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences

CC encode the novel proteins given in AAY11114 to AAY11367. The proteins,

CC isolated from Streptococcus pneumoniae, can be used in vaccines against

CC streptococcal infections and in assays for identifying compounds that

CC inhibit or activate the activity of the proteins. The antagonists can

CC be used to treat an individual having need to inhibit a bacterial

CC protein. Vectors expressing the proteins can be used to induce a

CC protective immune response in mammals.

SQ Sequence 192 AA;

Query Match 100.0%; Score 34; DB 18; Length 192;

Best Local Similarity 100.0%; Pred. No. 80;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
Db 91 hcleh 95

RESULT 5

ID AAB25440 standard; Protein; 229 AA.

XX AAB25440;

DT 27-NOV-2000 (first entry)

XX Pinus radiata cell signalling involved protein SEQ ID NO:759.

DE Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;

KM plant cell signalling; modulation; transgenic plant; pathogen; growth;

KW environmental change; development; cell proliferation; differentiation;

KW elongation; survival; disease resistance; nutrient metabolism.

XX Pinus radiata.

PN W0200042171-A1.

PD 20-JUL-2000.

PF 11-JAN-2000; 2000WO-US00724.

PR 12-JAN-1999; 99US-0228986.

PR 01-NOV-1999; 99US-0162866.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PI Strabala TJ, Nieuwenhuizen NJ;

PI WPI; 2000-476052/41.

PT Isolated polynucleotide encoding a polypeptide involved in cell

PT signalling used for generating transgenic plants with modified responses

PS Claim 3; Page 351-352; 527pp; English.

CC AAB79263 to AAB79736 and AAB25100 to AAB25570 represent polynucleotide

CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or

CC pine (Pinus radiata also known as Monterey pine). The protein sequences

CC are involved in cell signalling. The polynucleotide and protein

CC sequences can be used to modify the response of plant cells to external

CC signals e.g. environmental changes or pathogens during the growth and

CC development of a plant. They can be used to modify cell proliferation,

CC differentiation, elongation and survival, resistance to disease and

CC nutrient metabolism. Examples of modifications which can be produced are

CC altered fruit ripening and senescence of leaves and flowers e.g. to

CC delay senescence and prolong the life of cut flowers or enhance

CC senescence of reproductive organs to engineer sterile plants. Other

CC modifications can be used to delay senescence in selected cell types or

CC organs providing fruit and vegetables which have a longer shelf life

CC between harvest and consumption, or to decrease branching frequency in

CC forest tree species yielding long stretches of valuable knot-free clear

CC wood which can be used in solid timber furniture and veneers.

SQ Sequence 229 AA;

Query Match 100.0%; Score 34; DB 21; Length 229;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 124 hcleh 128

## RESULT 6

AAK47151  
 ID AAK47151 standard; Protein; 230 AA.

AC AAK47151;

DT 13-JUN-1994 (first entry)

DE IL-2 receptor gamma chain.

KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;

KW rheumatoid arthritis; transplant rejection; primer;

KW polymerase chain reaction; PCR; amplification.

OS Homo sapiens.

PN EP578932-A.

PD 19-JAN-1994.

PF 22-APR-1993; 93EP-0106561.

PR 23-APR-1992; 92JP-0104947.

PA (AJIN ) AJINOMOTO KK.

PA (SUGA/) SUGAMURA K.

PI Aseo H, Hamuro J, Nakamura M, Shlamamura T, Sugamura K;

PI Suzuki M, Takeshita T;

DR WPI: 1994-017546/03.

DR N-PSDB; AAQ54831.

PT DNA and protein sequences of IL-2 gamma chain - useful as immune

PT regulatory agents for treatment of e.g. rheumatoid arthritis and

PT transplant rejection

PS Disclosure: Page 22-23, 35-36; 50pp; English.

CC The human IL-2 receptor gamma chain preform (AAK47148), including the

CC signal peptide, is encoded by the sequence given in AAQ54828. The

CC mature protein (AAK47149) is encoded by sequence AAQ54829. A soluble

CC form of IL-2 receptor gamma chain (AAK47150) is encoded by AAQ54830.

CC While a soluble form suitable for expression in prokaryotes (AAK47151)

CC is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-

CC terminal sequence of IL-2 receptor gamma chain, and are used to

CC isolate IL2 receptor gamma chain receptor cDNA. Primers AAQ54826-27

CC are used to obtain the protein given in AAK47151.

CC

SO Sequence 230 AA;

Query Match 100.0%; Score 34; DB 15; Length 230;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 159 hcleh 163

## RESULT 7

AAK82934

ID AAK82934 standard; Protein; 230 AA.

AC AAK82934;

XX 26-FEB-1996 (first entry)

DT Interleukin 4 component common to the IL-2 receptor gamma chain.

DE Interleukin-4; IL-4; gamma chain component; immunosuppressants;

KW anti-allergy agent; signal transmission inhibitor; autoimmune;

KW disease; anti-inflammatory; anaphylactic shock; bronchial asthma;

KW Interleukin-2; IL-2; atopic dermatitis; urticaria.

OS Homo sapiens.

PN JP07149662-A.

PD 13-JUN-1995.

PF 07-SEP-1994; 94JP-0213706.

PR 08-SEP-1993; 93JP-0223574.

PA (AJIN ) AJINOMOTO KK.

PA (SUGA/) SUGAMURA K.

DR WPI: 1995-243601/32.

DR N-PSDB; AAT04952.

PT Novel interleukin-4 receptor monoclonal antibodies inhibit signal

PT transmission - useful as immunosuppressants and anti-allergy agents.

PS Example 1; Page 9; 11pp; Japanese.

CC AAT04952 encodes AAK82934 a component of the IL-4 receptor common to

CC the IL-2 receptor gamma chain molecule, which was used to generate

CC anti-IL-4 receptor monoclonal antibodies (mAbs). The mAbs (IL-4

CC signal transmission inhibitors) can be used as immunosuppressants

CC and anti-allergy agents, for the treatment of autoimmune and chronic

CC inflammatory diseases, e.g. anaphylactic shock, bronchial asthma,

CC atopic dermatitis and urticaria.

CC

SO Sequence 230 AA;

Query Match 100.0%; Score 34; DB 16; Length 230;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 160 hcleh 164

## RESULT 8

AAK47150

ID AAK47150 standard; Protein; 252 AA.

AC AAK47150;

DT 13-JUN-1994 (first entry)

DE IL-2 receptor gamma chain.

KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;

KW rheumatoid arthritis; transplant rejection; primer;

KW polymerase chain reaction; PCR; amplification; ss.

OS Homo sapiens.

PN Key

FT Peptide

FT 1..22

PN /Label= sig\_peptide

EP578932-A.

PD 19-JAN-1994.  
XX  
PF 22-APR-1993; 93EP-0106561.  
XX 23-APR-1992; 92JP-0104947.  
XX  
PA (AJIN ) AJINOMOTO KK.  
PA (SUGA/) SUGAMURA K.  
XX  
PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;  
PI Suzuki M, Takeshita T;  
XX  
DR WPI; 1994-017546/03.  
DR N-PSDB; AA054830.  
XX  
PT DNA and protein sequences of IL-2 gamma chain - useful as immune  
PT regulatory agents for treatment of e.g. rheumatoid arthritis and  
PT transplant rejection  
XX  
PS Disclosure; Page 21-22, 34-35; 50pp; English.  
XX  
CC The human IL-2 receptor gamma chain preform (AAR47148), including the  
CC signal peptide, is encoded by the sequence given in AA054828. The  
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble  
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,  
CC while a soluble form suitable for expression in prokaryotes (AAR47151)  
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-  
CC terminal sequence of IL-2 receptor gamma chain, and are used to  
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27  
CC are used to obtain the protein given in AAR47151.  
XX  
SQ Sequence 252 AA;  
  
Query Match 100.0%; Score 34; DB 15; Length 252;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 HCLEH 5  
Db 181 hcleh 185  
  
RESULT 9  
AA054733  
ID AA054733 standard; Protein: 270 AA.  
XX  
AC AA054733;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 69867.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139494.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144332.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0145388.  
PR 03-AUG-1999; 99US-0145389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158237.  
PR 12-OCT-1999; 99US-0158366.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 34; DB 21; Length 270;  
Best Local Similarity 100.0%; Pred. No. 1; 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLFH 5  
Db 187 hclfh 191

RESULT 10  
AAV81769  
ID AAV81769 standard; Protein; 297 AA.  
XX  
AC AAV81769;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID127.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
kidney disease; diabetes; immunosuppressive disorder; otitis media;  
pneumococcal septicemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB02452.  
XX  
PR 27-JUL-1998; 98GB-0016336.  
PR 19-MAR-1999; 99US-0125329.  
XX  
PA (MICR-) MICROBIAL TECHNICS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
XX  
DR N-PSDB; AA291865.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
treatment and prophylaxis of bacterial infections  
XX  
PS Claim 2; Page 63; 76pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis.

SQ Sequence 297 AA;

QY 1 HCLEH 5  
| | | | |  
Db 214 hcleh 218

Query Match 100.0%; Score 34; DB 21; Length 297;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
AAR47149  
ID AAR47149 standard; Protein; 347 AA.  
XX AAR47149;  
AC AAR47149;  
DT 13-JUN-1994 (first entry)  
XX  
DE IL-2 receptor gamma chain.  
XX  
KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;  
KW rheumatoid arthritis; transplant rejection; primer;  
KW polymerase chain reaction; PCR; amplification.  
XX  
OS Homo sapiens.  
XX  
PN EP578932-A.  
XX  
PD 19-JAN-1994.  
XX  
PF 22-APR-1993; 93EP-0106561.  
XX  
PR 23-APR-1992; 92JP-0104947.  
XX  
PA (AJIN ) AJINOMOTO K.  
PA (SUGA) SUGAMURA K.  
XX  
PI Aseo H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;  
PI Suzuki M, Takeshita T;  
XX  
DR WPI; 1994-017546/03.  
DR N-PSDB; AA054829.  
XX  
PT DNA and protein sequences of IL-2 gamma chain - useful as immune  
PT regulatory agents for treatment of e.g. rheumatoid arthritis and  
PT transplant rejection  
XX  
PS Claim 4; Page 41; 50pp; English.  
XX  
CC The human IL-2 receptor gamma chain preform (AAR47148), including the  
CC signal peptide, is encoded by the sequence given in AA054828. The  
CC mature protein (AAR47149) is encoded by sequence AA054829. A soluble  
CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AA054830,

CC while a soluble form suitable for expression in prokaryotes (AAR47151)  
CC is encoded by AA054831. Primers 1-6 (AA054820-25) are based on the N-  
CC terminal sequence of IL-2 receptor gamma chain, and are used to  
CC isolate IL2 receptor gamma chain receptor cDNA. Primers AA054826-27  
CC are used to obtain the protein given in AAR47151.  
XX  
SQ Sequence 347 AA;

QY 1 HCLEH 5  
| | | | |  
Db 159 hcleh 163

Query Match 100.0%; Score 34; DB 15; Length 347;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12  
AAB71682  
ID AAB71682 standard; protein; 363 AA.  
XX  
AC AAB71682;  
DT 10-MAY-2001 (first entry)  
XX  
DE Bos taurus Interleukin-2 receptor gamma.  
XX  
KW Cytokine receptor common gamma chain like; CRGCL; human;  
KW tumours; infections; inflammatory; immune disorder;  
KW neurodegenerative; cardiovascular; disorder.  
XX  
OS Bos taurus.  
XX  
PN WO200112672-A2.  
XX  
PD 22-FEB-2001.  
XX  
PF 17-AUG-2000; 2000WO-US22493.  
XX  
PR 18-AUG-1999; 99US-0376430.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SA, Rosen CA, Moore PA;  
XX  
DR WPI; 2001-147547/15.  
XX  
PT New nucleic acid molecule encoding a human cytokine receptor common  
PT gamma chain like polypeptide, useful for treating, preventing and/or  
PT diagnosing e.g. tumors, inflammatory diseases and immunodeficiency  
PT conditions -  
XX  
PS Disclosure; Fig 2; 288pp; English.  
XX  
CC The present invention relates to a human cytokine receptor  
CC common gamma chain like protein (CRGCL). The invention is useful  
CC for treating, preventing and/or diagnosing conditions such  
CC as tumours, infections, inflammatory diseases,  
CC immune disorders, neurodegenerative disorder and cardiovascular  
CC disorder.  
XX  
SQ Sequence 363 AA;

QY 1 HCLEH 5  
| | | | |  
Db 172 hcleh 176

Query Match 100.0%; Score 34; DB 22; Length 363;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13  
 AAR47148  
 ID AAR47148 standard; Protein: 369 AA.  
 XX  
 AC AAR47148;  
 XX  
 DT 13-JUN-1994 (first entry)  
 XX  
 DE IL-2 receptor gamma chain.  
 XX  
 KW Interleukin-2 receptor gamma chain; IL-2; receptor; immunoregulator;  
 KW rheumatoid arthritis; transplant rejection; primer; PCR;  
 KW polymerase chain reaction; amplification; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /Label= Sig\_peptide  
 XX  
 PN EP578932-A.  
 XX  
 PD 19-JAN-1994.  
 XX  
 PE 22-APR-1993; 93EP-0106561.  
 XX  
 PR 23-APR-1992; 92JP-0104947.  
 XX  
 PA (AJIN ) AJINOMOTO KK.  
 PA (SUGA ) SUGAMURA K.  
 XX  
 PI Asao H, Hamuro J, Nakamura M, Shimamura T, Sugamura K;  
 PI Suzuki M, Takeshita T;  
 DR N-PSDB: AAQ54828.  
 DR WPI: 1994-017546/03.  
 XX  
 PT DNA and protein sequences of IL-2 gamma chain - useful as immune  
 PT regulatory agents for treatment of e.g. rheumatoid arthritis and  
 PT transplant rejection  
 XX  
 PS Disclosure: Page 16-17, 29-30; 50pp; English.  
 XX  
 CC The human IL-2 receptor gamma chain preform (AAR47148), including the  
 CC signal peptide, is encoded by the sequence given in AAQ54828. The  
 CC mature protein (AAR47149) is encoded by sequence AAQ54829. A soluble  
 CC form of IL-2 receptor gamma chain (AAR47150) is encoded by AAQ54830.  
 CC While a soluble form suitable for expression in prokaryotes (AAR47151)  
 CC is encoded by AAQ54831. Primers 1-6 (AAQ54820-25) are based on the N-  
 CC terminal sequence of IL-2 receptor gamma chain, and are used to  
 CC isolate IL2 receptor gamma chain, receptor cDNA. Primers AAQ54826-27  
 CC are used to obtain the protein given in AAR47151.  
 XX  
 SQ Sequence 369 AA:  
 Query Match 100.0%; Score 34; DB 15; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HCLEH 5  
 DB 181 hcleh 185  
 RESULT 14  
 AAW31646  
 ID AAW31646 standard; Protein: 482 AA.  
 XX  
 AC AAW31646;  
 XX  
 DT 21-MAY-1998 (first entry)

XX  
 DE Human cytokine receptor gc chain-ig fusion protein.  
 XX  
 KW Blocking agent; monoclonal antibody; CP:88; immunological disease;  
 KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;  
 KW insulin-dependent diabetes; inflammatory bowel disease;  
 KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;  
 KW graft versus host disease; psoriasis; immunosuppressive; therapy.  
 XX  
 OS Chimeric - Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT Protein 1..254  
 FT /note="gc chain N-terminal region"  
 FT Protein 255..482  
 FT /note="IgG1 constant region"  
 FT Region 255..264  
 FT /note="IgG1 hinge region"  
 FT Domain 264..482  
 FT /note="IgG1 CH2 and CH3 constant domains1"  
 XX  
 PN WO9743416-A1.  
 XX  
 PD 20-NOV-1997.  
 XX  
 PE 09-MAY-1997; 97WO-US07870.  
 XX  
 PR 10-MAY-1996; 96US-0017466.  
 XX  
 PA (BIOJ ) BIOGEN INC.  
 XX  
 PI Benjamin CD, Burkly LC, Hession C, Whitty A;  
 PI WPI: 1998-008885/01.  
 DR N-PSDB: AAT97439.  
 DR  
 XX  
 PT Blocking agents of the gamma common chain of cytokine receptors -  
 PT particularly monoclonal antibodies, used to induce T cell anergy for  
 PT treatment of immunological diseases  
 XX  
 PS Example 1; Page 79-80; 11pp; English.  
 XX  
 CC This polypeptide comprises a fusion between the N-terminal 254  
 CC amino acids of the human mature cytokine receptor gamma common (gc)  
 CC chain and the hinge region and CH2 and CH3 constant domains of  
 CC human IgG1. The fusion was expressed from clone PLB001 (see  
 CC AAT97439) in COS-7 cells, and used to generate murine anti-human gc  
 CC specific monoclonal antibodies (MAbs), including CP.B8 produced by  
 CC hybridoma ATCC HB 12107. The invention provides compositions and  
 CC methods for inhibiting cytokine signalling using gc chain blocking  
 CC agents for the treatment of immunological diseases such as  
 CC myasthenia gravis, rheumatoid arthritis, lupus, multiple sclerosis,  
 CC insulin-dependent diabetes, inflammatory bowel disease, sympathetic  
 CC ophthalmia, uveitis, allergy, asthma, parasitic infection, graft  
 CC vs. host disease or psoriasis. A preferred gc blocking agent is  
 CC Mab CP.B8 or its Fab fragment (see also AAW31647-48).  
 XX  
 SQ Sequence 482 AA:  
 Query Match 100.0%; Score 34; DB 19; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HCLEH 5  
 DB 181 hcleh 185  
 RESULT 15  
 AAY92202  
 ID AAY92202 standard; Protein: 691 AA.



Search completed: August 20, 2002, 11:07:18  
 Job time: 6103 sec

XX AAY92202;  
 AC  
 XX  
 DT 01-AUG-2000 (first entry)  
 XX  
 DE Fusion polypeptide 603, IL-4 trap.  
 XX  
 KM IL-4 trap; cytokine; antagonist; CNTF; receptor; fusion protein;  
 XX cytostatic; immunomodulator; osteopathic.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO200018932-A2.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 22-SEP-1999; 99WO-US22045.  
 XX  
 PR 25-SEP-1998; 98US-0101858.  
 PR 19-MAY-1999; 99US-0313942.  
 XX  
 PA (REG-) REGENERON PHARM INC.  
 XX  
 PI Stahl N, Yancopoulos GD;  
 XX  
 DR WPI: 2000-293165/25.  
 DR N-PSDB; AAA09044.  
 XX  
 PT Isolated nucleic acid molecule for treating cytokine-related diseases  
 PT or disorders encodes a fusion polypeptide capable of binding a cytokine  
 PT to form a nonfunctional complex  
 XX  
 PS Example 6; Fig 22A-D; 152pp; English.  
 CC This sequence shows fusion polypeptide 603, which is capable of  
 CC binding cytokine IL-4 to form a non-functional complex.  
 CC The invention concerns production of antagonists to any cytokine that  
 CC utilizes an alpha specificity determining component, which when combined  
 CC with the cytokine, binds to a first beta signal transducing component to  
 CC form a non-functional intermediate which then binds to a second beta  
 CC signal transducing component causing beta-receptor dimerization, the  
 CC soluble alpha specificity determining component of the receptor  
 CC (SR-alpha) and the extracellular domain of the first beta signal  
 CC transducing component of the cytokine receptor (beta-1) are combined to  
 CC form heterodimers (SR-alpha:beta-1) that act as antagonist to the  
 CC cytokine by binding the cytokine to form a non-functional complex. The  
 CC receptor components are shared by cytokines such as the CNTF (ciliary  
 CC neurotrophic factor) family of cytokines. The invention provides the  
 CC basis for the development of IL-6 antagonists, as they show that if, in  
 CC the presence of a ligand, a non-functional intermediate complex,  
 CC consisting of the ligand, its alpha receptor and its beta-1 receptor  
 CC component, can be formed, it will effectively block the action of the  
 CC ligand. Effective antagonists of IL-6 or CNTF consist of heterodimers  
 CC of the extracellular domains of the alpha specificity determining  
 CC components of their receptors and the extracellular domain of gp130.  
 CC The resultant heterodimers, function as high-affinity traps, rendering  
 CC the cytokine inaccessible to form a signal transducing complex with the  
 CC native membrane-bound forms of their receptor. The nucleic acids and  
 CC polypeptides are useful for treating cytokine-related diseases or  
 CC disorders such as osteoporosis and primary and secondary effects of  
 CC cancer including multiple myeloma or cachexia.  
 XX  
 SO Sequence 691 AA;

Query Match 100.0%; Score 34; DB 21; Length 691;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
 |||||  
 Db 181 hcleh 185

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:37 ; Search time 100.84 Seconds  
(without alignments)  
4.764 Million cell updates/sec

Title: US-09-824-286-16

Perfect score: 34

Sequence: 1 HCLEH 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR\_71:\*  
2: pir1:\*  
3: pir2:\*  
4: pir3:\*  
5: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	129	2	AB1993
2	34	100.0	287	2	A95129
3	34	100.0	297	2	F97999
4	34	100.0	351	1	D84430
5	34	100.0	364	1	D86281
6	34	100.0	369	2	A42565
7	34	100.0	373	2	A55718
8	34	100.0	1044	2	T00342
9	34	100.0	2117	2	T36180
10	32	94.1	85	2	F82600
11	32	94.1	1829	2	T34239
12	31	91.2	30	2	JC1360
13	31	91.2	161	2	F72647
14	31	91.2	191	2	T23594
15	31	91.2	304	1	C84430
16	31	91.2	333	1	H86381
17	31	91.2	417	2	I50407
18	31	91.2	425	2	A46251
19	31	91.2	477	1	C2C2AM
20	31	91.2	515	2	B96825
21	31	91.2	632	2	D84921
22	31	91.2	694	2	G97400
23	31	91.2	746	2	AG2618
24	31	91.2	1314	2	G02870
25	31	91.2	2078	2	T25400
26	30	88.2	186	2	A11986
27	30	88.2	249	2	T46184
28	30	88.2	335	2	E87625
29	30	88.2	436	2	JC1497

30	29	85.3	75	2	D27393	11K inner spore co
31	29	85.3	119	2	AD3319	hypothetical prote
32	29	85.3	140	2	T16574	hypothetical prote
33	29	85.3	155	2	B95086	hypothetical prote
34	29	85.3	155	2	F97953	cytidine deaminase
35	29	85.3	165	2	F82136	conserved hypothet
36	29	85.3	186	2	AE1622	B. subtilis COMEB
37	29	85.3	187	2	AC1260	hypothetical prote
38	29	85.3	187	2	S75522	hypothetical prote
39	29	85.3	242	2	A96538	hypothetical prote
40	29	85.3	278	2	D70036	exopolysaccharide
41	29	85.3	299	2	S31771	HRP11 protein - ma
42	29	85.3	301	2	S31782	HRP11 protein - ma
43	29	85.3	301	2	S31773	HRP11 protein - ma
44	29	85.3	308	2	T29317	hypothetical prote
45	29	85.3	309	2	A25942	histidine/alanine-

## ALIGNMENTS

RESULT 1  
AB1993  
virulence associated protein C [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AB1993  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Irig  
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-Fixing Cyanobacterium  
A:Reference number: AB1807, MID:21595285, PMID:11759840  
A:Accession: AB1993  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-129 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA077861.1; PID:g17135316; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: vapC  
C:Superfamily: virulence-associated protein vapC

Query Match 100.0%; Score 34; DB 2; Length 129;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
|||||  
Db 103 HCLEH 107

RESULT 2  
A95129  
transcription regulator MutR, probable [imported] - Streptococcus pneumoniae (strain  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: A95129  
R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; I  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MID:21357209; PMID:11463916  
A:Accession: A95129  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-287 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75226.1; PID:g14972591; GSPDB:GN00164; TIGR:  
A:Experimental source: strain TIGR4  
C:Genetics:

A:Gene: SP115

Query Match 100.0%; Score 34; DB 2; Length 287;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
DB 204 HCLEH 208

RESULT 3  
F97999

transcription regulator rgpd [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
R:Accession: F97999  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Ee, R.; Leblang, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Nye, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: F97999  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-297 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99826.1; PID:G15458639; GSPDB:GN00174  
C:Genetics:  
A:Gene: rgpd

Query Match 100.0%; Score 34; DB 2; Length 297;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
DB 214 HCLEH 218

RESULT 4

D84430  
probable acid phosphatase (EC 3.1.3.2) At2g01890 precursor [similarity] - Arabidopsis th  
N:Alternate names: purple acid phosphatase  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Nov-2001  
C:Accession: D84430  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: D84430  
A:Molecule type: DNA  
A:Residues: 1-351 <STO>  
A:Cross-references: GB:AE002093; NID:94522012; PIDN:AAD21785.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g01890  
A:Map position: 2  
C:Superfamily: tartrate-resistant acid phosphatase; phosphoric monoester hydrolase  
C:Keywords: iron; metalloprotein; phosphoric monoester hydrolase  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-351/Product: acid phosphatase #status predicted <MAT>  
F:46-125/Domain: phosphoesterase core homology <PEC>  
F:52,65,88,270/Binding site: iron (Asp, Asp, Tyr, His) #status predicted  
F:85,123,233,268/Binding site: iron (Asp, Asn, His, His) #status predicted  
F:124,242/Active site: His #status predicted

Query Match 100.0%; Score 34; DB 1; Length 351;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
DB 270 HCLEH 274

RESULT 5

D86281  
probable acid phosphatase (EC 3.1.3.2) F10B6.10 precursor [similarity] - Arabidopsis  
N:Alternate names: purple acid phosphatase  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
R:Accession: D86281  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maltl, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: D86281  
A:Molecule type: DNA  
A:Residues: 1-364 <STO>  
A:Cross-references: GB:AE005172; NID:98778212; PIDN:AAK79221.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F10B6.10  
A:Map position: 1  
C:Superfamily: tartrate-resistant acid phosphatase; phosphoesterase core homology  
C:Keywords: iron; metalloprotein; phosphoric monoester hydrolase  
F:1-12/Domain: signal sequence #status predicted <SIG>  
F:13-364/Product: acid phosphatase #status predicted <MAT>  
F:75-152/Domain: phosphoesterase core homology <PEC>  
F:81,114,117,281/Binding site: iron (Asp, Asp, Tyr, His) #status predicted  
F:114,150,244,279/Binding site: iron (Asp, Asn, His, His) #status predicted  
F:151,253/Active site: His #status predicted

Query Match 100.0%; Score 34; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
DB 281 HCLEH 285

RESULT 6

A42565  
interleukin-2 receptor gamma chain - human  
C:Species: Homo sapiens (man)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
R:Accession: A42565; A46591; I54332  
R:Takeshita, T.; Asao, H.; Ohtani, K.; Ishii, N.; Kumaki, S.; Tanaka, N.; Munakata, H  
Science 257, 379-382, 1992  
A:Title: Cloning of the gamma chain of the human IL-2 receptor.  
A:Reference number: A42565; MUID:92335883  
A:Accession: A42565  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid; protein  
A:Residues: 1-369 <TRK>  
A:Cross-references: GB:D11086; NID:9303611; PIDN:BA01857.1; PID:9219890  
A:Experimental source: MOLT beta lymphoid cells  
A:Note: sequence extracted from NCBI backbone (NCBI:109167)  
R:Noguchi, M.; Adelstein, S.; Cao, X.; Leonard, W.J.  
J. Biol. Chem. 268, 13601-13608, 1993  
A:Title: Characterization of the human Interleukin-2 receptor gamma chain gene.

A:Reference number: A46591; MUID:93293887  
A:Accession: A46591  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-369 <RES>  
A:Cross-references: GB:U12183; NID:9307056; PIDN:AA459145.1; PID:9307058  
R:Puck, J.M.; Deschenes, S.M.; Porter, J.C.; Dutra, A.S.; Brown, C.J.; Willard, H.F.; He Hum. Mol. Genet. 2, 1099-1104, 1993  
A:Title: The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked  
A:Reference number: I54332; MUID:94004847  
A:Accession: I54332  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-369 <RE2>  
A:Cross-references: GB:U19546; NID:9349631; PIDN:AAC37524.1; PID:9349632  
C:Genetics:  
A:Gene: GDB:IL2RG; SCIDX1; IMD4  
A:Cross-references: GDB:U134807; OMIM:308380  
A:Map position: Xq13.1-Xq13.1  
A:Introns: 39/1; 90/2; 152/1; 198/3; 253/1; 285/2; 308/3  
A:Note: defects are associated with an X-linked form of severe combined immunodeficiency  
C:Superfamily: interleukin-2 receptor gamma chain  
C:Keywords: cytokine receptor; duplication; immunodeficiency; severe combined immunodeficiency

Query Match 100.0%; Score 34; DB 2; Length 369;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
Db 181 HCLEH 185

RESULT 7  
A55718  
Interleukin-2 receptor gamma chain precursor - dog  
C:Species: Canis lupus familiaris (dog)  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 23-Jul-1999  
C:Accession: A55718  
R:Henthorn, P.S.; Somberg, R.L.; Fimlant, V.M.; Puck, J.M.; Patterson, D.F.; Felsburg, F Genomics 23, 69-74, 1994  
A:Title: IL-2Rgamma gene microdeletion demonstrates that canine X-linked severe combined  
A:Reference number: A55718; MUID:95130114  
A:Accession: A55718  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-373 <HEN>  
A:Cross-references: GB:U04361; NID:9517411; PIDN:AAC48403.1; PID:9517412  
C:Superfamily: interleukin-2 receptor gamma chain  
C:Keywords: cytokine receptor; duplication

Query Match 100.0%; Score 34; DB 2; Length 373;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
Db 181 HCLEH 185

RESULT 8  
T00342  
hypothetical protein KIA0580 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00342  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O. DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete  
A:Reference number: Z14086; MUID:98290545  
A:Accession: T00342

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1044 <NAG>  
A:Cross-references: EMBL:AB011152; NID:93043683; PIDN:BA25506.1; PID:93043684  
A:Experimental source: brain; clone HD0601  
C:Genetics:  
A:Note: KIA0580

Query Match 100.0%; Score 34; DB 2; Length 1044;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
Db 954 HCLEH 958

RESULT 9  
T36180  
CDA peptide synthetase III - Streptomyces coelicolor (fragment)  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 01-Dec-2000  
C:Accession: T36180  
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream submitted to the EMBL data library, March 1999  
A:Reference number: Z21600  
A:Accession: T36180  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2117 <SAU>  
A:Cross-references: EMBL:AL035707; PIDN:CAB38876.1; GSPDB:GN00070; SCOREDB:SCE63.01  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: caaP3; SCOREDB:SCE63.01  
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology  
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F:208-662/Domain: acetate-CoA ligase homology <ACLI>  
F:678-746/Domain: acyl carrier protein homology <ACPI>  
F:1282-1725/Domain: acetate-CoA ligase homology <ACI2>  
F:1740-1806/Domain: acyl carrier protein homology <ACR2>  
F:1710-1772/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 100.0%; Score 34; DB 2; Length 2117;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
|||||  
Db 466 HCLEH 470

RESULT 10  
F82600  
hypothetical protein XP2097 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: F82600  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: F82600  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-85 <SIM>  
A:Cross-references: GB:AE004025; GB:AE003849; NID:99107217; PIDN:AAF84896.1; GSPDB:G A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R Bioness, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000  
A:Authors: Perrella, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kuramae, E.E.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B  
A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa, J.V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
A:Genetics:  
A:Gene: XF2097

Query Match 94.1%; Score 32; DB 2; Length 85;  
Best Local Similarity 80.0%; Pred. No. 21;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HCLEH 5  
Db 15 HCLDH 19

RESULT 11  
T34239  
hypothetical protein F26F12.7 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34239  
R:Wilson, R.; Bentley, D.; Gatlung, S.  
Submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid F26F12.  
A:Reference number: Z21493  
A:Accession: T34239  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1829 <MIL>  
A:Cross-references: EMBL:U55373; PIDN:AAC25894.1; GSPDB:GND0023; CESP:F26F12.7  
A:Experimental source: strain Bristol N2; clone F26F12  
C:Genetics:  
A:Gene: CESP:F26F12.7  
A:Map position: 5  
A:Introns: 110/3; 441/3; 801/2; 1244/3; 1693/2; 1784/1

Query Match 94.1%; Score 32; DB 2; Length 1829;  
Best Local Similarity 80.0%; Pred. No. 3,1e+02;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HCLEH 5  
Db 299 HCLDH 303

RESULT 12  
JC1360  
hypothetical 3k protein (HV3 5' region) - human  
C:Species: *Homo sapiens* (man)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 28-May-1999  
C:Accession: JC1360  
R:Shimoto, T.; Tsuchimochi, H.; McGregor, C.G.A.; Mutch, H.; Shimizu, T.; Kurachi, Y.  
Biochem. Biophys. Res. Commun. 189, 617-624, 1992  
A:Title: Molecular cloning and characterization of the platelet-activating factor recept  
A:Reference number: JC1359; MUID:93112021  
A:Accession: JC1360  
A:Molecule type: mRNA  
A:Residues: 1-30 <SUG>  
A:Cross-references: GB:S52624; NID:g262469; PIDN:AAB24696.1; PID:g1680456  
A:Experimental source: heart

Query Match 91.2%; Score 31; DB 2; Length 30;

Best Local Similarity 80.0%; Pred. No. 13;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HCLEH 5  
Db 14 HCLDH 18

RESULT 13  
F72647  
hypothetical protein APE0612 - *Aeropyrum pernix* (strain K1)  
C:Species: *Aeropyrum pernix*  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: F72647  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aero*  
A:Reference number: A72450; MUID:99310339  
A:Accession: F72647  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-161 <KAW>  
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BA79582.1; PID:g5104267  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0612  
C:Superfamily: *Aeropyrum pernix* hypothetical protein APE0612

Query Match 91.2%; Score 31; DB 2; Length 161;  
Best Local Similarity 80.0%; Pred. No. 56;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HCLEH 5  
Db 144 HCLDH 148

RESULT 14  
T23594  
hypothetical protein K10H10.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Sep-2000  
C:Accession: T23594  
R:Percy, C.  
Submitted to the EMBL Data Library, December 1996  
A:Reference number: Z19766  
A:Accession: T23594  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-191 <MIL>  
A:Cross-references: EMBL:Z83236; PIDN:CAB05781.1; GSPDB:GND0020; CESP:K10H10.5  
A:Experimental source: clone K10H10  
C:Genetics:  
A:Gene: CESP:K10H10.5  
A:Map position: 2  
A:Introns: 22/2; 89/1  
C:Superfamily: *Caenorhabditis elegans* hypothetical protein K10H10.4

Query Match 91.2%; Score 31; DB 2; Length 191;  
Best Local Similarity 80.0%; Pred. No. 65;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
Db 41 HCLDH 45

RESULT 15  
C84430  
probable acid phosphatase (EC 3.1.3.2) At2g01880 precursor [similarity] - *Arabidopsis*

N;Alternate names: purple acid phosphatase  
 C;Species: Arabidopsis thaliana (mouse-ear cross)  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Nov-2001  
 C;Accession: C84430  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A;Reference number: AB4420; MUID:20083487  
 A;Accession: C84430  
 A;Molecule type: DNA  
 A;Residues: 1-304 <STO>  
 A;Cross-references: GB:AE002093; NID:94522007; PIDN:AAD21780.1; GSPDB:GN00139  
 C;Genetics:  
 A;Gene: At2g01880  
 A;Map position: 2  
 C;Superfamily: tartrate-resistant acid phosphatase; phosphoesterase core homology  
 C;Keywords: iron; metalloprotein; phosphoric monoester hydrolase  
 F;1-21/Domain: signal sequence #status predicted <SIG>  
 F;22-304/Product: acid phosphatase #status predicted <MAT>  
 F;42-121/Domain: phosphoesterase core homology <PEC>  
 F;48,81,84,226/Binding site: iron (Asp, Asp, Tyr, His) #status predicted  
 F;81,119,189,224/Binding site: iron (Asp, Asn, His, His) #status predicted  
 F;120,198/Active site: His #status predicted

Query Match 91.2% Score 31; DB 1; Length 304;  
 Best Local Similarity 80.0% Pred. No. 99;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 HCLEH 5  
 |||:|  
 Db 226 HCLQH 230

Search completed: August 20, 2002, 11:10:39  
 Job time: 5669 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:02 : Search time 55.29 Seconds

(without alignments)  
3.501 Million cell updates/sec

Title: US-09-824-286-16

Perfect score: 34

Sequence: 1 HCLEH 5

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	369	1	CYRG_HUMAN
2	34	100.0	373	1	CYRG_CANFA
3	34	100.0	379	1	CYRG_BOVIN
4	31	91.2	220	1	GT29_FASHE
5	31	91.2	416	1	PGK_CHICK
6	31	91.2	425	1	ADFP_MOUSE
7	31	91.2	477	1	GUNA_CLOTT
8	31	91.2	1314	1	Y197_HUMAN
9	31	91.2	1402	1	AFAM_MOUSE
10	30	88.2	75	1	COTD_BACSU
11	29	85.3	312	1	CDAL_YEAST
12	29	85.3	314	1	MHPB_ECOLI
13	29	85.3	326	1	INH5_ECOLI
14	29	85.3	332	1	HRP1_PLAFA
15	29	85.3	338	1	INSR_ECOLI
16	29	85.3	361	1	COL9_ARATH
17	29	85.3	450	1	ADFP_BOVIN
18	29	85.3	714	1	GREM_CHLAM
19	29	85.3	716	1	BGAL_THETU
20	29	85.3	742	1	ZFA_MOUSE
21	29	85.3	791	1	G6PE_HUMAN
22	29	85.3	805	1	DE19_GAEEL
23	29	85.3	870	1	POL_ISR
24	29	85.3	1313	1	MIP1_SCHPO
25	29	85.3	32	1	Y180_TREPA
26	28	82.4	61	1	TBA4_MAIZE
27	28	82.4	158	1	NUDB_HAEIN
28	28	82.4	189	1	REX3_MOUSE
29	28	82.4	242	1	PEPE_XENLA
30	28	82.4	323	1	RFC4_YEAST
31	28	82.4	360	1	CB2R_HUMAN
32	28	82.4	418	1	TBA_AJECA
33	28	82.4	418	1	TBA_AJECA

34	28	82.4	444	1	TBA_ONCKE	P30436 oncorhynch
35	28	82.4	447	1	TBA2_ELEIN	O22348 eleusine in
36	28	82.4	447	1	TBA_AVECA	P34690 caenorhabdi
37	28	82.4	448	1	TBA2_CABEL	P05215 homo sapien
38	28	82.4	448	1	TBA4_HUMAN	P09644 gallus gall
39	28	82.4	449	1	TBA5_CHICK	P24633 emericella
40	28	82.4	449	1	TBA1_EMENT	P53372 pneumocysti
41	28	82.4	449	1	TBA1_PNECA	P06604 drosophila
42	28	82.4	449	1	TBA2_DROME	P04689 schizosacch
43	28	82.4	449	1	TBA2_SCHPO	P05218 mus musculu
44	28	82.4	449	1	TBA6_MOUSE	P11480 physarum po
45	28	82.4	449	1	TBAE_PHYPO	

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	369 AA.
CYRG_HUMAN				
ID	CYRG_HUMAN			
AC	P31785			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cytokine receptor common gamma chain precursor (Gamma-C) (Interleukin-2 receptor gamma chain) (IL-2R gamma chain) (p64) (CD132 antigen).			
GN	IL2RG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
XP	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RA	MEDLINE=9235883; PubMed=1631559;			
RA	Takeshita T., Asao H., Ohtani K., Ishii N., Kumaki S., Tanaka N.,			
RT	"Cloning of the gamma chain of the human IL-2 receptor.";			
RL	Science 257:379-382(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=9329387; PubMed=8514792;			
RA	Noguchi M., Adelstein S., Cao X., Leonard W.J.;			
RT	"Characterization of the human Interleukin-2 receptor gamma chain			
RL	gene.";			
RL	J. Biol. Chem. 268:13601-13608(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS ASP-114 AND ASN-153.			
RX	MEDLINE=94004847; PubMed=8401490;			
RA	Puck J.M., Deschenes S.M., Porter J.C., Dutra A.S., Brown C.J.,			
RT	Willard H., Henchorn P.S.;			
RL	"The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated			
RL	in X-linked severe combined immunodeficiency, SCIDX1.";			
RN	Hum. Mol. Genet. 2:1099-1104(1993).			
RP	[4]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94090315; PubMed=8266076;			
RA	Kondo M., Takeshita T., Ishii N., Nakamura M., Watanabe S.,			
RT	Arai K.-I., Sugamura K.;			
RL	"Sharing of the interleukin-2 (IL-2) receptor gamma chain between			
RL	receptors for IL-2 and IL-4.";			
RN	Science 262:1874-1877(1993).			
RP	[5]			
RP	IDENTIFICATION AS A IL-4R SUBUNIT.			
RX	MEDLINE=94090317; PubMed=8266078;			
RA	Russell S.M., Kegan A.D., Harada N., Nakamura Y., Noguchi M.,			
RT	Leonard W.J.,			
RL	"Interleukin-2 receptor gamma chain: a functional component of the			
RL	interleukin-4 receptor.";			
RL	Science 262:1880-1883(1993).			
RN	[6]			
RP	IDENTIFICATION AS A IL-7R SUBUNIT.			

RA MEDLINE-94090316; PubMed-8266077;  
RX Noouchi M., Nakamura Y., Russell S.M., Ziegler S.F., Tsang M., Cao X.  
RA Leonard W.J., "Interleukin-7 receptor gamma chain: a functional component of the  
RT Interleukin-7 receptor."  
RL Science 262:1877-1880(1993).  
RN [7]  
RP 3D-STRUCTURE MODELING OF 57-248.  
RX MEDLINE-9511955; PubMed-7529123;  
RA Bamorough P., Hedgecock C.J., Richards W.G.;  
RT "The interleukin-2 and Interleukin-4 receptors studied by molecular  
RT modelling."  
RL Structure 2:839-851(1994).  
RN [8]  
RP VARIANTS XSCID PHE-115; CYS-240 AND ILE-241.  
RX MEDLINE-94130970; PubMed-8296698;  
RA Disanto J.P., Dautry-Varsat A., Certain S., Fischer A.,  
RA de Saint Basile G.;  
RT "Interleukin-2 (IL-2) receptor gamma chain mutations in X-linked  
RT severe combined immunodeficiency disease result in the loss of  
RT high-affinity IL-2 receptor binding."  
RL Eur. J. Immunol. 24:475-479(1994).  
RN [9]  
RP VARIANT XSCID LYS-68.  
RX MEDLINE-94375038; PubMed-8088810;  
RA Markiewicz S., Subtil A., Dautry-Varsat A., Fischer A.,  
RA de Saint Basile G.;  
RT "Detection of three nonsense mutations and one missense mutation in  
RT the Interleukin-2 receptor gamma chain gene in SCIDX1 that  
RT differentially affect the mRNA processing."  
RL Genomics 21:291-293(1994).  
RN [10]  
RP VARIANT XSCID HIS-162.  
RX MEDLINE-94300093; PubMed-8027558;  
RA Ishii N., Asao H., Kimura Y., Takeshita T., Nakamura M., Tsuchiya S.,  
RA Kono T., Meeda M., Uchiyama T., Sugamura K.;  
RT "Impairment of ligand binding and growth signaling of mutant IL-2  
RT receptor gamma-chains in patients with X-linked severe combined  
RT immunodeficiency."  
RL J. Immunol. 153:1310-1317(1994).  
RN [11]  
RP VARIANT XSCID ASN-39.  
RX MEDLINE-95023932; PubMed-7937790;  
RA Disanto J.P., Rieux-Laucat F., Dautry-Varsat A., Fischer A.,  
RA de Saint Basile G.;  
RT "Defective human Interleukin 2 receptor gamma chain in an atypical X  
RT chromosome-linked severe combined immunodeficiency with peripheral T  
RT cells."  
RL Proc. Natl. Acad. Sci. U.S.A. 91:9466-9470(1994).  
RN [12]  
RP VARIANTS XSCID CYS-226 AND HIS-226.  
RX MEDLINE-95397841; PubMed-7668284;  
RA Peppet A.E., Buckley R.H., Small T.N., Puck J.M.;  
RT "Two mutational hotspots in the Interleukin-2 receptor gamma chain  
RT gene causing human X-linked severe combined immunodeficiency."  
RL Am. J. Hum. Genet. 57:564-571(1995).  
RN [13]  
RP VARIANT XSCID SFR-183.  
RX MEDLINE-96013903; PubMed-7557965;  
RA Clark P.A., Lester T., Genet S., Jones A.M., Hendriks R.,  
RA LeVinsky R.L., Kinon C.;  
RT "Screening for mutations causing X-linked severe combined  
RT immunodeficiency in the IL-2R gamma chain gene by single-strand  
RT conformation polymorphism analysis."  
RL Hum. Genet. 96:427-432(1995).  
RN [14]  
RP VARIANT XSCID GLN-237 G-H-W INS.  
RX MEDLINE-95164726; PubMed-7860773;  
RA Puck J.M., Peppet A.E., Bedard P.-M., Laframboise R.;  
RT "Female germ line mosaicism as the origin of a unique IL-2 receptor  
RT gamma-chain mutation causing X-linked severe combined  
RT immunodeficiency."  
RL J. Clin. Invest. 95:895-899(1995).

[15]  
 MEDLINE-95190013; PubMed-7883965;  
 RA Schmalstieg F.C., Leonard W.J., Noguchi M., Berg M., Rudloff H.E.,  
 RX Denney R.M., Dave S.K., Brooks E.G., Goldman A.S.;  
 RT "Missense mutation in exon 7 of the common gamma chain gene causes a  
 RL moderate form of x-linked combined immunodeficiency."; J.  
 J. Clin. Invest. 95:1169-1173(1995).  
 [16]  
 VARIANT XSCID ARG-115.  
 RX MEDLINE-97042245; PubMed-8900089;  
 RA Stephan V., Wahn V., Le Deist F., Dirksen U., Broeker B.,  
 RA Mueller-Fleckenstein I., Horneff G., Schroten H., Fischer A.,  
 RA de Saint Basile G.;  
 RT "Atypical x-linked severe combined immunodeficiency due to possible  
 RL spontaneous reversion of the genetic defect in T cells."; New  
 Engl. J. Med. 335:1563-1567(1996).  
 [17]  
 VARIANT XSCID GLN-285.  
 RX MEDLINE-97295088; PubMed-9150740;  
 RA Jones A.M., Clark P.A., Katz F., Genet S., McMahon C., Alterman L.,  
 RA Cant A., Kinon C.;  
 RT "B-cell-negative severe combined immunodeficiency associated with a  
 RL common gamma chain mutation."; Hum. Genet. 99:677-680(1997).  
 [18]  
 VARIANT XSCID CYS-222.  
 RX MEDLINE-98064061; PubMed-9399950;  
 RA Sharfe N., Shahar M., Rolfman C.M.;  
 RT "An interleukin-2 receptor gamma chain mutation with normal thymus  
 morphology."; Clin. Invest. 100:3036-3043(1997).  
 RL  
 J. Clin. Invest. 100:3036-3043(1997).  
 CC - FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF  
 CC INTERLEUKINS.  
 CC - SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND  
 CC PROBABLY ALSO THE IL-13 RECEPTOR.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A SEVERE COMBINED  
 CC IMMUNODEFICIENCY, WHICH IS KNOWN AS AGAMMAGLOBULINEMIA, SWISS TYPE  
 CC OR X-LINKED SEVERE COMBINED IMMUNODEFICIENCY (XSCID).  
 CC - SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC - SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC - DATABASE: NAME=PROT; NOTE=CD guide CD132 entry;  
 CC MMW="http://www.ncbi.nlm.nih.gov/prov/cd/cd132.htm".  
 CC - DATABASE: NAME=IL2RGbase; NOTE=X-linked SCID mutation database;  
 CC WWW="http://www.nhgri.nih.gov/DIR/GMBS/SCID/".  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).  
 -----  
 CC  
 DR EMBL: D11086; BAA01857.1; -;  
 DR EMBL: L12183; AAA59145.1; -;  
 DR EMBL: L12178; AAA59145.1; JOINED.  
 DR EMBL: L12176; AAA59145.1; JOINED.  
 DR EMBL: L12177; AAA59145.1; JOINED.  
 DR EMBL: L12179; AAA59145.1; JOINED.  
 DR EMBL: L12180; AAA59145.1; JOINED.  
 DR EMBL: L12181; AAA59145.1; JOINED.  
 DR EMBL: L12182; AAA59145.1; JOINED.  
 DR EMBL: L19546; AAC37524.1; -;  
 DR PIR: A42565; A42565.  
 DR PDB: 1ILM; 26-JAN-95.  
 DR MIM: 308380; -;  
 DR MIM: 300400; -;  
 DR InterPro: IPR002996; CRA.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003531; Hematopo\_receptor\_S\_F1.

Query Match 100.0%; Score 34; DB 1; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 8.8;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 181 HCLEH 185

```

RESULT 2
CYRG_CANFA STANDARD; PRT; 373 AA.
ID CYRG_CANFA
AC P40321;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (gamma-C)
DE (interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN IL2RG.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95130114; PubMed=7829104;
RA Henthorn P.S., Somberg R.L., Fimant V.M., Puck J.M., Patterson D.F.,
RA Felsburg P.J.;
RT "IL-2R gamma gene microdeletion demonstrates that canine X-linked
RT severe combined immunodeficiency is a homologue of the human
RT disease.";
RL Genomics 23:69-74(1994).
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: DEFECTS IN IL2RG ARE THE CAUSE OF A CANINE X-LINKED
CC SEVERE COMBINED IMMUNODEFICIENCY.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U04361; AAC48403.1; -.
CC HSSP: P31785; 11LN.
CC InterPro: IPR002996; CRIA.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR003531; Hematopo_receptor_S_F1.
CC Pfam: PF00041; fn3; 1.
CC SMART: SM00060; FN3; 1.
CC PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
CC Receptor; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 22
CC CHAIN 1 22
CC CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC FIBRONECTIN TYPE-III.
CC DOMAIN 151 249
CC TRANSSEM 284 373
CC DOMAIN 284 373
CC DISULFID 62 72
CC POTENTIAL.
CC CARBOHYD 102 115
CC DISULFID 24 24
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 71 71
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 75 75
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 84 84
  
```

FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 373 AA; 42516 MW; 03A0DE1F8B089DBB CRC64;

Query Match 100.0%; Score 34; DB 1; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 8.9;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 181 HCLEH 185

```

RESULT 3
CYRG_BOVIN STANDARD; PRT; 379 AA.
ID CYRG_BOVIN
AC O95118;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytokine receptor common gamma chain precursor (gamma-C)
DE (interleukin-2 receptor gamma chain) (IL-2R gamma chain) (P64).
GN IL2RG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96268473; PubMed=8672241;
RA Yeo J., Stone R.T., Solinas-Toldo S., Fries R., Beattie C.W.;
RA "Cloning and chromosomal mapping of bovine interleukin-2 receptor
RA gamma gene.";
RL DNA Cell Biol. 15:453-459(1996).
CC -1- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
CC INTERLEUKINS.
CC -1- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
CC PROBABLY ALSO THE IL-13 RECEPTORS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U33748; AAB07812.1; -.
CC HSSP: P31785; 11LN.
CC InterPro: IPR002996; CRIA.
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR003531; Hematopo_receptor_S_F1.
CC Pfam: PF00041; fn3; 1.
CC SMART: SM00060; FN3; 1.
CC PROSITE: PS01355; HEMATOPO_REC_S_F1; 1.
CC Receptor; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 22
CC CHAIN 1 22
CC CYTOKINE RECEPTOR COMMON GAMMA CHAIN.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC FIBRONECTIN TYPE-III.
CC DOMAIN 270 290
CC TRANSSEM 291 379
CC DOMAIN 291 379
CC DISULFID 68 78
CC POTENTIAL.
CC CARBOHYD 109 122
CC DISULFID 77 77
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 81 81
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 90 90
  
```

FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 379 AA; 43037 MM; 33CFAD9C9B032178 CRC64;

Query Match 100.0%; Score 34; DB 1; Length 379;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 11111  
 DB 188 HCLDH 192

RESULT 4  
 GT29\_FASHE STANDARD; PRT; 220 AA.  
 ID GT29\_FASHE  
 AC P56598;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutathione S-transferase 26 kDa 1 (EC 2.5.1.18) (GST1) (FHL) (GST  
 DE class-alpha).  
 OS Fasciola hepatica (Liver fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;  
 OC Rhabditiophora; Eulicthiophora; Revertospermatia; Medifusata;  
 OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;  
 OC Fasciolidae; Fasciolidae; Fasciola.  
 OC NCBI\_TaxID=6192;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92153506; PubMed=1740183;  
 RA Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;  
 RT "Molecular characterization of cDNA sequences encoding glutathione S-  
 RL transfeases of Fasciola hepatica."  
 RL Exp. Parasitol. 74:232-237(1992).  
 RN [2]  
 RP ERRATUM.  
 RX MEDLINE=94039664; PubMed=8224094;  
 RA Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;  
 RL Exp. Parasitol. 77:385-386(1993).  
 RN [3]  
 RP SEQUENCE OF 22-220 FROM N.A.  
 RA Crameri S.;  
 RL Patent number WO9008819, 09-AUG-1990.  
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER  
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.  
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE  
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO  
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF  
 CC HEMATIN IN THE PARASITE GUT.  
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: A00993; CA00118.1; -;  
 DR HSSP: P31670; IPHE.  
 DR InterPro: IPR004046; GST\_C.  
 DR InterPro: IPR004045; GST\_N.  
 DR Pfam: PF00043; GST\_C; 1.  
 DR Pfam: PF02798; GST\_N; 1.  
 KW Transferase; Antigen; Multigene family.  
 FT INIT MET 0 BY SIMILARITY.  
 FT CONFLICT 22 22 Y->V (IN REF. 3).  
 FT CONFLICT 110 111 DP->VS (IN REF. 3).

FT CONFLICT 189 189 A->P (IN REF. 3).  
 SQ SEQUENCE 220 AA; 25598 MM; 27B9F150B75D101F CRC64;

Query Match 91.2%; Score 31; DB 1; Length 220;  
 Best Local Similarity 80.0%; Pred. No. 20;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 11111  
 DB 166 HCLDH 170

RESULT 5  
 PGK\_CHICK STANDARD; PRT; 416 AA.  
 ID PGK\_CHICK  
 AC P51903;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Phosphoglycerate kinase (EC 2.7.2.3).  
 GN PGK.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=94238148; PubMed=8182283;  
 RA Rauen K.A., Le Ciel C.D., Abbott U.K., Hutchison N.J.;  
 RT "Localization of the chicken PGK gene to chromosome 4p by  
 RL fluorescence in situ hybridization."  
 RL J. Hered. 85:147-150(1994).  
 CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate -> ADP + 3-  
 CC phospho-D-glyceroyl phosphate.  
 CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L37101; AAC42219.1; -;  
 DR HSSP: P00560; 10PG.  
 DR InterPro: IPR001576; PGK.  
 DR Pfam: PF00162; PGK; 1.  
 DR PRINTS: PR00477; PHGLYCKINASE.  
 DR PROSITE: PS00111; PGLYCERATE\_KINASE; 1.  
 KW Transferase; Kinase; Glycolysis.  
 FT INIT MET 0 BY SIMILARITY.  
 SQ SEQUENCE 416 AA; 44585 MM; 216A7A298590C31A CRC64;

Query Match 91.2%; Score 31; DB 1; Length 416;  
 Best Local Similarity 80.0%; Pred. No. 37;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 11111  
 DB 48 HCLDH 52

RESULT 6  
 ADP\_MOUSE STANDARD; PRT; 425 AA.  
 ID ADP\_MOUSE  
 AC P43883;

DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Adipophilin (Adipose differentiation-related protein) (ADRP).  
GN ADRP OR ADRP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adipocyte;  
RX MEDLINE=92390349; PubMed=1518805;  
RA Jiang H.P., Serrero G.;  
RT "Isolation and characterization of a full-length cDNA coding for an  
RT adipose differentiation-related protein";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:7856-7860(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H; TISSUE=Adipose tissue;  
RX MEDLINE=93315154; PubMed=8325636;  
RA Eisinger D.P., Serrero G.;  
RT "Structure of the gene encoding mouse adipose differentiation-related  
RT protein (ADRP).";  
RL Genomics 16:638-644(1993).  
CC -1- FUNCTION: MAY BE INVOLVED IN DEVELOPMENT AND MAINTENANCE OF  
CC ADIPOSE TISSUE.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated;  
CC -1- TISSUE SPECIFICITY: ADIPOSE TISSUE SPECIFIC. EXPRESSED ABUNDANTLY  
CC AND PREFERENTIALLY IN FAT PADS.  
CC -1- INDUCTION: BY DEXAMETHASONE.  
CC -1- SIMILARITY: BELONGS TO THE PERILIPIN FAMILY.  
CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M93275; AAA37176.1; -;  
DR EMBL: L09734; -; NOT\_ANNOTATED\_CDS.  
DR MGD: MGI:87920; Adfp.  
DR InterPro: IPR004279; perilipin.  
DR Pfam: PF03036; perilipin; 1.  
KW Membrane.  
SQ SEQUENCE 425 AA; 46664 MW; 82624E6CE3429C22 CRC64;  
Query Match 91.2%; Score 31; DB 1; Length 425;  
Best Local Similarity 80.0%; Pred. No. 38;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HCLEH 5  
Db 299 HCVH 303

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB\_10682;  
RX MEDLINE=85157393; PubMed=3980433;  
RA Beguin P., Cornet P., Aubert J.-P.;  
RT "Sequence of a cellulase gene of the thermophilic bacterium  
RT Clostridium thermocellum".;  
RL J. Bacteriol. 162:102-105(1985).  
RN [2]  
RP SEQUENCE OF 33-46.  
RC STRAIN=JW20;  
RX MEDLINE=96267082; PubMed=8664281;  
RA Choi S.K., Ljungdahl L.G.;  
RT "Dissociation of the cellulose of Clostridium thermocellum in the  
RT presence of ethylenediaminetetraacetic acid occurs with the formation  
RT of truncated polypeptides.";  
RL Biochemistry 35:4897-4905(1996).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).  
RX MEDLINE=96398613; PubMed=8805535;  
RA Alzari P.M., Soucoun H., Dominguez R.;  
RT "The crystal structure of endoglucanase CelA, a family 8 glycosyl  
RT hydrolase from Clostridium thermocellum.";  
RL Structure 4:265-275(1996).  
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
CC GLUCANS.  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose.  
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL  
CC HYDROLASES).  
CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: K03088; AAA83521.1; -;  
DR PIR: A23100; CZCLAM.  
DR PDB: ICEM; 1J-JAN-97.  
DR InterPro: IPR002105; Dockerin\_1.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR002037; Glyco\_hydro\_8.  
DR Pfam: PF00404; Dockerin\_1; 2.  
DR Pfam: PF01270; Glyco\_hydro\_8; 1.  
DR PRINTS: PR00735; GLHYDRASE8.  
DR PROSITE: PS00018; EF\_HAND; UNKNOWN\_2.  
DR PROSITE: PS00448; CLOS\_CELLULOSE\_RPT; 2.  
DR PROSITE: PS00812; GLYCOSYL\_HYDROL\_F8; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal;  
KW 3D-structure.  
FT SIGNAL 1 32  
FT CHAIN 33 477  
FT ACT\_SITE 95 95  
FT ACT\_SITE 152 152  
FT DOMAIN 417 472  
FT REPEAT 417 440  
FT REPEAT 449 472  
FT REPEAT 41 41  
FT CONFLICT 41 41  
SQ SEQUENCE 477 AA; 52594 MW; BA5A0AD5022E8A51 CRC64;  
Query Match 91.2%; Score 31; DB 1; Length 477;  
Best Local Similarity 80.0%; Pred. No. 42;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HCLEH 5

```

Db      190 HCVH 194
      ||:|
      11:11

RESULT      8
ID_Y197_HUMAN  STANDARD;  PRT;  1314 AA.
AC      Q12769;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protein KIAA0197 (Fragment).
GN      KIAA0197.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Bone marrow;
RX      MEDLINE=96281124; PubMed=8724849;
RA      Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT      "Prediction of the coding sequences of unidentified human genes. V.
RT      The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT      analysis of cDNA clones from human cell line Kg-1.";
RL      DNA Res. 3:17-24(1996).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D83781; BAA12110.1; -.
FT      NON TER
SQ      SEQUENCE      1314 AA; 148828 MW; B2FF622FF6A6AA0 CRC64;

Query Match      91.2%; Score 31; DB 1; Length 1314;
Best Local Similarity      80.0%; Pred. No. 1.1e+02;
Matches      4; Conservative      1; Mismatches      0; Indels      0; Gaps      0;

QY      1 HCLEH 5
      ||:|
      11:11
Db      255 HCVH 259

RESULT      9
ID_Y197_MOUSE  STANDARD;  PRT;  1402 AA.
AC      O940M3;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protein KIAA0197 (grl-13).
GN      KIAA0197 OR GT11-13.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/SVJ;
RA      Van de Putte T., Cozijnsen M., Dewulf N., Tylzanowski P., Lonnoy O.,
RA      Huylenbroeck D.;
RT      "Mus musculus mRNA for grl-13 (gene trap locus-13), similar to human
RT      KIAA0197 gene (883781), complete cds.";
RL      Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

[illegible]

FT DISULFID 340 385 BY SIMILARITY.  
 FT DISULFID 384 393 BY SIMILARITY.  
 FT DISULFID 416 462 BY SIMILARITY.  
 FT DISULFID 461 470 BY SIMILARITY.  
 FT DISULFID 483 499 BY SIMILARITY.  
 FT DISULFID 498 509 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 611 AA; 69635 MW; 39E46B6E723F89C8 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 611;  
 Best Local Similarity 80.0%; Pred. No. 83;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 Db 508 HCFEH 512

RESULT 11  
 COTD\_BACSU STANDARD; PRT; 75 AA.  
 AC P07791;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Spore coat protein D.  
 GN COTD.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88011308; PubMed=2821284;  
 RA Donovan W., Zheng L., Sandman K., Losick R.;  
 RT "Genes encoding spore coat polypeptides from Bacillus subtilis."  
 RL J. Mol. Biol. 196;1-10(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / MARBURG;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,  
 RA Seror P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 the sera and kgd loci cloned in a yeast artificial chromosome."  
 RL Microbiology 142:2005-2016(1996).  
 RN [3]  
 RP SEQUENCE OF 1-6 FROM N.A.  
 RX MEDLINE=90230303; PubMed=1691789;  
 RA Zheng L., Losick R.;  
 RT "Cascade regulation of spore coat gene expression in Bacillus  
 subtilis."  
 RL J. Mol. Biol. 212:645-660(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X05681; CAA29168.1; ALT\_TERM.  
 DR EMBL; L47838; AAB38470.1; -;  
 DR EMBL; Z99115; CAB14137.1; -;  
 DR PIR; D27393; D27393.  
 DR Subtilisin; Bg10493; cotd.

KW Sporulation; Complete proteome.  
 SQ SEQUENCE 75 AA; 8840 MW; A5019889CA6CC0EA CRC64;

Query Match 85.3%; Score 29; DB 1; Length 75;  
 Best Local Similarity 80.0%; Pred. No. 18;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 Db 17 HCFEH 21

RESULT 12  
 CDAL\_YEAST STANDARD; PRT; 312 AA.  
 ID CDAL\_YEAST  
 AC Q06703;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Chitin deacetylase 1 precursor (EC 3.5.1.41).  
 GN CDAL OR YLR308W OR L2142.1.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Maridis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifken L., Riles L., Taich A., Trevasaks E., Vignati D.,  
 RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97279228; PubMed=9133736;  
 RA Mishra C.A., Semino C.E., McCreath K.J., de la Vega H., Jones B.J.,  
 RA Specht C.A., Robbins P.W.;  
 RT "Cloning and expression of two chitin deacetylase genes of  
 Saccharomyces cerevisiae."  
 RL Yeast 13:327-336(1997).  
 CC -----  
 CC -1- FUNCTION: HYDROLYZES THE N-ACETAMIDO GROUPS OF N-ACETYL-D-  
 CC GLUCOSAMINE RESIDUES IN CHITIN.  
 CC -1- CATALYTIC ACTIVITY: Chitin + H(2)O = chitosan + acetate.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY DURING SPOULATION.  
 CC -1- SIMILARITY: TO OTHER POLYSACCHARIDE DEACETYLASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U17247; AAB67355.1; -;  
 DR SGD; S0004298; CDAL.  
 DR InterPro; IPR002509; Polysac\_deacet.  
 DR Pfam; PF01522; Polysac\_deacet.1.  
 KW Hydrolyase; Chitin degradation; Signal; Glycoprotein; Sporulation.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 312 CHITIN DEACETYLASE 1.  
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 312 AA; 35693 MW; 4033EEFA803F6AF1 CRC64;

Query Match 85.3%, Score 29; DB 1; Length 312;  
 Best Local Similarity 80.0%; Pred. NO. 68;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 HCLEH 5  
 1111  
 Db 107 HCTEH 111

RESULT 13  
 MHPB\_ECOLI STANDARD; PRT: 314 AA.  
 ID MHPB\_ECOLI P54711: P77461; P77048;  
 AC P54711: P77461; P77048;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE 2,3-dihydroxyphenylpyruvate 1,2-dioxygenase (EC 1.13.11.-)  
 GN MHPB OR B0348 OR Z0446 OR ECS0403.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / W3110;  
 RA Spence E.L., Kawamukai M., Sanvoisin J., Braven H., Bugl T.D.H.;  
 RT "Catechol dioxygenases from Escherichia coli (Mhp) and Alcaligenes  
 RT eutrophus (Mpci): sequence analysis and biochemical properties of a  
 RT third family of extradiol dioxygenases."; J. Bacteriol. 178:5249-5256(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / CS520;  
 RA Ferrandez A., Garcia J.L., Diaz E.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
 RA Nemat A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RA MEDLINE=21074933; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lin A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."; Nature 409:529-533(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / RIMD 0509952;  
 RA MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kunara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12."; DNA Res. 8:11-22(2001).  
 RN [7]  
 RP SEQUENCE OF 1-15, AND CHARACTERIZATION.  
 RX MEDLINE=94002258; PubMed=8393388;  
 RA Bugl T.D.H.;  
 RT "Overproduction, purification and properties of 2,3-  
 RT dihydroxyphenylpyruvate 1,2-dioxygenase from Escherichia coli."; Biochim. Biophys. Acta 1207:258-264(1993).  
 RL Biochim. Biophys. Acta 1207:258-264(1993).  
 CC -1- FUNCTION: EXTRADIOL CLEAVAGE OF 2,3-DIHYDROXYPHENYLPROPYONIC ACID.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- PATHWAY: 3-HYDROXYPHENYLPROPYONATE DEGRADATION PATHWAY.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SIMILARITY: TO A.EUTROPHUS MCP1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D86239; BAA13053.1; -;  
 DR EMBL: Y09555; CAAT70748.1; -;  
 DR EMBL: AE000142; AAC73451.1; -;  
 DR EMBL: U73857; AAB18072.1; -;  
 DR EMBL: AE005214; AAC54699.1; -;  
 DR EMBL: AP002551; BAB33826.1; -;  
 DR Ecogene: EG20274; mhpB.  
 DR InterPro: IPR004183; L19B.  
 DR Pfam: PF02900; L19B.  
 KW Oxidoreductase; Dioxygenase; Aromatic hydrocarbons catabolism; Iron;  
 KW Repeat: Complete proteome.  
 FT DOMAIN 1 171 1.  
 FT 172 314 2.  
 FT CONFLICT 138 140 ING -> NKA (IN REF. 1).  
 FT CONFLICT 152 152 R -> H (IN REF. 1).  
 FT CONFLICT 157 157 A -> T (IN REF. 1).  
 SQ SEQUENCE 314 AA; 34196 MW; ELD5A8574E5DFE05 CMC64;

Query Match 85.3%, Score 29; DB 1; Length 314;  
 Best Local Similarity 80.0%; Pred. NO. 69;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 HCLEH 5  
 1111  
 Db 6 HCLSH 10

RESULT 14  
 INHS\_ECOLI STANDARD; PRT: 326 AA.  
 ID INHS\_ECOLI P76071;  
 AC P76071;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Transposase *insH* for insertion sequence element ISSY.  
 GN INSH OR B1370.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,



RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 MAU B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -1- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION  
 CC SEQUENCE IS5.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSPOSASE FAMILY 15.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000234; AAC74452.1; -  
 DR EcoGene; EG40008; insH.  
 DR InterPro: IPR002559; Transposase\_11.  
 DR Pfam: PF01609; Transposase\_11.  
 KM Transposable element; Transposition; DNA-binding; DNA recombination;  
 KW Complete proteome.  
 SQ SEQUENCE 326 AA; 37777 MW; 1A3C80D7FD5D8B5B CRC64;

Query Match Best Local Similarity 85.3%; Score 29; DB 1; Length 326;  
 Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HCLEH 5  
 ||:|  
 Db 64 HCMQH 68

RESULT 15  
 ID HRP1\_PLAFA STANDARD; PRT; 332 AA.  
 AC P05227;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 13-DEC-1998 (Rel. 37, Last annotation update)  
 DE Histidine-rich protein precursor (Clone PHRP-II).  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=86287395; PubMed=3016741;  
 RX Wellens T.E., Howard R.J.;  
 RT "Homologous genes encode two distinct histidine-rich proteins in a  
 RT cloned isolate of Plasmodium falciparum.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6065-6069(1986).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M13986; AAA51639.1; -  
 DR EMBL: K03509; AAA51639.1; JOINED.  
 KM Malaria; Repeat; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 332 HISTIDINE-RICH PROTEIN.  
 SQ SEQUENCE 332 AA; 35128 MW; 330842B99C5FCF5B CRC64;

Query Match Best Local Similarity 85.3%; Score 29; DB 1; Length 332;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HCLEH 5  
 ||:|  
 Db 328 HCLRH 332

Search completed: August 20, 2002, 11:33:03  
 Job time: 1452 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:56 ; Search time 191.14 Seconds  
(without alignments)  
4.525 Million cell updates/sec

Title: US-09-824-286-16  
Perfect score: 34  
Sequence: 1 HCLEH 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhcc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_rvivirus:\*  
17: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	172	16	09CM60 pasteurella
2	34	100.0	287	16	097QT4 streptococc
3	34	100.0	314	10	09SD28
4	34	100.0	331	10	09SL79
5	34	100.0	332	10	09L180
6	34	100.0	351	10	09S1S5
7	34	100.0	364	10	09L0W0
8	34	100.0	375	13	09PDC9
9	34	100.0	701	4	096MA4
10	34	100.0	891	2	09JL38
11	34	100.0	1044	4	09Y4E4
12	34	100.0	1704	4	096L70
13	34	100.0	2117	2	09Z4Z9
14	32	94.1	85	16	09PBP3
15	32	94.1	294	2	09JNR7
16	32	94.1	317	13	09DGR3

17	32	94.1	364	5	09VDM7	09vdm7 drosophila
18	32	94.1	459	10	09FVK3	09fvk3 glycine max
19	32	94.1	1829	5	019815	019815 caenorhabd
20	31	91.2	30	4	09P2Z8	09p2z8 homo sapien
21	31	91.2	103	12	09Y7G4	09y7g4 bovine aden
22	31	91.2	161	17	09YEG4	09yeg4 aetopyrum p
23	31	91.2	174	2	09AD00	09ad00 streptomyce
24	31	91.2	191	5	045682	045682 caenorhabd
25	31	91.2	228	16	09CK00	09ck00 pasteurella
26	31	91.2	252	16	0984F0	0984f0 rhizobium l
27	31	91.2	257	2	093SN7	093sn7 prochloroco
28	31	91.2	304	10	09S1S6	09s1s6 arabidopsis
29	31	91.2	333	10	09FRR2	09frr2 arabidopsis
30	31	91.2	338	10	09SCX8	09scx8 arabidopsis
31	31	91.2	425	11	09CY19	09cy19 mus musculu
32	31	91.2	515	10	09SAK7	09sak7 arabidopsis
33	31	91.2	607	10	09PEP9	09pep9 arabidopsis
34	31	91.2	632	10	082261	082261 arabidopsis
35	31	91.2	633	2	09F392	09f392 streptomyce
36	31	91.2	748	11	099MW0	099mw0 mus musculu
37	31	91.2	827	5	09U6M1	09u6m1 leishmania
38	31	91.2	1497	5	095YH6	095yh6 drosophila
39	31	91.2	1571	5	093H7	093h7 drosophila
40	31	91.2	1612	5	09VE37	09ve37 drosophila
41	31	91.2	2078	5	P91834	P91834 caenorhabd
42	31	91.2	2454	3	09UVP2	09uvp2 emericella
43	31	91.2	2454	3	09UV56	09uv56 emericella
44	30	88.2	168	12	091MP9	091mp9 lumpy skin
45	30	88.2	175	2	044138	044138 anabena sp

## ALIGNMENTS

RESULT 1  
ID 09CM60 PRELIMINARY; PRT; 172 AA.  
AC 09CM60;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE NTPA.  
GN NTPA OR PM0981.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=2145866; PubMed=11248100;  
RA May B.J., Zhang O., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida Fm70.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
DR EMBL; AEO06137; AKR03065.1; -;  
DR InterPro; IPR000086; NUDIX\_hydrolase.  
DR PROSITE; PS00893; NUDIX; 1.  
KW Complete proteome.  
SQ SEQUENCE 172 AA; 20168 MW; F8FCB0B3E0B71F53 CRC64;

Query Match 100.0%; Score 34; DB 16; Length 172;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
|||||  
Db 93 HCLEH 97

RESULT 2  
097QT4 PRELIMINARY; PRT; 287 AA.  
ID 097QT4

AC 097074;  
DT 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE TRANSCRIPTIONAL REGULATOR MUTR, PUTATIVE.  
GN SP115.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TIGR4;  
RA MEDLINE=21357209; PubMed=11463916;  
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E., Khouiri H., Wolf A.M., Uterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldlyum T.V., Anguoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae";  
RL Science 293:498-506(2001);  
DR EMBL: AE007412; AKR75226.1; -  
DR TIGR: SP115; -  
DR InterPro: IPR001387; HTH\_3.  
DR InterPro: IPR003975; Shal\_channel.  
DR Pfam: PF01381; HTH\_3; 1.  
DR PRINTS: PR01497; SHALCHANNEL.  
KW Complete proteome.  
SQ SEQUENCE 287 AA; 34077 MW; B88A5C58FF6FE976 CRC64;  
  
Query Match 100.0%; Score 34; DB 16; Length 287;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HCLEH 5  
Db 204 HCLEH 208  
  
RESULT 3  
Q9SLD8 PRELIMINARY; PRT; 314 AA.  
AC Q9SD28;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PUTATIVE PURPLE ACID PHOSPHATASE PRECURSOR.  
GN PAP.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20314483; PubMed=10854785;  
RA Schenk G., Guddat L.W., Ge Y., Carrington L.E., Hume D.A.,  
RA Hamilton S., de Jersey J.;  
RT "Identification of mammalian-like purple acid phosphatases in a wide  
RT range of plants";  
RL Gene 250:117-125(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20517299; PubMed=11062342;  
RA Oddie G.W., Schenk G., Angel N.Z., Walsh N., Guddat L.W.,  
RA de Jersey J., Cassidy A.I., Hamilton S.E., Hume D.A.;  
RT "Structure, function, and regulation of tartrate-resistant acid  
RT phosphatase";

RL Bone 27:575-584(2000).  
CC -I CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O - AN  
CC ALCOHOL + ORTHOPHOSPHATE.  
DR EMBL: AF200827; AAF19823.1; -  
DR InterPro: IPR003266; PA\_phosphatase.  
DR InterPro: IPR009334; Ser\_thr\_phosphatase.  
DR Pfam: PF02227; PA\_phosphatase; 1.  
KW Hydrolyase; Iron; Signal; Zinc.  
FT SIGNAL 1 30  
FT CHAIN 31 314 POTENTIAL.  
FT SEQUENCE 314 AA; 35933 MW; 6982EBB4DEAC03CF CRC64;  
  
Query Match 100.0%; Score 34; DB 10; Length 314;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HCLEH 5  
Db 254 HCLEH 258  
  
RESULT 4  
Q9LL79 PRELIMINARY; PRT; 331 AA.  
AC Q9LL79;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PUTATIVE PURPLE ACID PHOSPHATASE.  
GN PAP.  
OS Phaseolus vulgaris (kidney bean) (French bean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
CN PAP.  
RT "Identification of mammalian-like purple acid phosphatases in a wide  
RT range of plants";  
RL Gene 250:117-125(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. RUFUS;  
RX MEDLINE=20314483; PubMed=10854785;  
RA Schenk G., Guddat L.W., Ge Y., Carrington L.E., Hume D.A.,  
RA Hamilton S., de Jersey J.;  
RT "Identification of mammalian-like purple acid phosphatases in a wide  
RT range of plants";  
RL Gene 250:117-125(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. RUFUS;  
RX MEDLINE=20517299; PubMed=11062342;  
RA Oddie G.W., Schenk G., Angel N.Z., Walsh N., Guddat L.W.,  
RA de Jersey J., Cassidy A.I., Hamilton S.E., Hume D.A.;  
RT "Structure, function, and regulation of tartrate-resistant acid  
RT phosphatase";  
RL Bone 27:575-584(2000).  
DR EMBL: AF236109; AAF60317.1; -  
DR InterPro: IPR005058; Peptidase\_S26.  
DR InterPro: IPR009334; Ser\_thr\_phosphatase.  
DR PROSITE: PS00761; SPASE\_I\_3; UNKNOWN\_1.  
SQ SEQUENCE 331 AA; 37652 MW; 662A8A2D9EF3788 CRC64;  
  
Query Match 100.0%; Score 34; DB 10; Length 331;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HCLEH 5  
Db 244 HCLEH 248  
  
RESULT 5  
Q9LL80 PRELIMINARY; PRT; 332 AA.

AC 09L180;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE PURPLE ACID PHOSPHATASE.  
 GN PAP.  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OX NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20314483; PubMed=10854785;  
 RA Schenk G., Guddat L.W., Ge Y., Carrington L.E., Hume D.A.,  
 RA Hamilton S., de Jersey J.;  
 RT "Identification of mammalian-like purple acid phosphatases in a wide  
 RT range of plants.";  
 RL Gene 250:117-125(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20517299; PubMed=11062342;  
 RA Oddie G.W., Schenk G., Angel N.Z., Walsh N., Guddat L.W.,  
 RA de Jersey J., Cassidy A.I., Hamilton S.E., Hume D.A.;  
 RT "Structure, function, and regulation of tartrate-resistant acid  
 RT phosphatase.";  
 RL Bone 27:575-584(2000).  
 DR EMBL: AF236108; AAF60316.1; -;  
 DR InterPro: IPR000934; Ser\_thr.phosphatse.  
 SQ INTERPRO: 332 AA; 37764 MW; B9E1463A946D9B49 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 252 HCLEH 256

RESULT 6  
 Q9S1S5 PRELIMINARY; PRT; 351 AA.  
 AC 09S1S5;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE PURPLE ACID PHOSPHATASE.  
 GN AT2G01890.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umeyan L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Coppenhaver G.P., Preuss D., Nierman W.C., Venter J.C.;  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=CV. COLUMBIA;  
 RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O - AN  
 CC ALCOHOL + ORTHOPHOSPHATE.  
 DR EMBL: AC007069; AAD21785.1; -;  
 DR InterPro: IPR003266; PA\_phosphatase.  
 DR Pfam: PF02227; PA\_phosphatase; 1.  
 DR HydroLase; Iron; Zinc.  
 KW HydroLase; Iron; Zinc.  
 SQ SEQUENCE 351 AA; 39998 MW; 335136FA234E8916 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 270 HCLEH 274

RESULT 7  
 Q9LOW0 PRELIMINARY; PRT; 364 AA.  
 AC 09LOW0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE F10B6.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,  
 RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,  
 RA Hansen N.F., Huizar L., Kremetska I., Lenz C., Li J., Liu S.,  
 RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaya V., Yu G.,  
 RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome  
 RT 1.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiu J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC006917; AAF79221.1; -;  
 DR InterPro: IPR000934; Ser\_thr.phosphatse.  
 SQ SEQUENCE 364 AA; 41839 MW; 192AA2367701CEC0 CRC64;

Query Match 100.0%; Score 34; DB 10; Length 364;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HCLEH 5  
 |||||  
 Db 281 HCLEH 285

RESULT 8  
Q9DDC9 PRELIMINARY; PRT; 375 AA.  
ID Q9DDC9  
AC Q9DDC9  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CHICKEN GONADOTROPIN-RELEASING HORMONE RECEPTOR.  
GN CGNRH-R.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21269321; PubMed=11112780;  
RA Sun Y.M., Flanagan C.A., Illing N., Ott T.R., Sellar R., Fromme B.J.,  
Hapgood J., Sharp P., Sealton S.C., Millar R.P.;  
RT "A chicken gonadotropin-releasing hormone receptor that confers  
agonist activity to mammalian antagonists: identification of D-Lys6 in  
the ligand and extracellular loop two of the receptor as  
determinants.";  
RT J. Biol. Chem. 276:7754-7761(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Sun Y.M.;  
RL Thesis (1998), Department of Chemical Pathology,  
University of Cape Town, Cape Town, South Africa.  
DR EMBL: AJ304414; CAC18674.1;-  
DR InterPro: IPR000276; GPCR\_Rhodopsn.  
DR Pfam: PF00001; 7tm\_1; 1.  
DR PRINTS: PR00237; GPCR\_RHODOPSN.  
DR PROSITE: PS00237; G\_PROTEIN\_RECEP\_F1\_1; UNKNOWN\_1.  
DR PROSITE: PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
KW Receptor.  
SQ SEQUENCE 375 AA; 42097 MW; 7DD866F0C6B95CE4 CRC64;

Query Match Best Local Similarity 100.0%; Score 34; DB 13; Length 375;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HCLEH 5  
Db 28 HCLEH 32

RESULT 9  
Q96MA4 PRELIMINARY; PRT; 701 AA.  
ID Q96MA4  
AC Q96MA4  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CDNA FLJ32706 FTS; CLONE TEST1200627 (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
Holtuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
Yamanoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;

RT "NEDO human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK057268; BAB71404.1;-  
FT NON\_TER  
SQ SEQUENCE 701 AA; 79822 MW; 99F46B28BBD7BCA1 CRC64;

Query Match Best Local Similarity 100.0%; Score 34; DB 4; Length 701;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HCLEH 5  
Db 517 HCLEH 521

RESULT 10  
Q93L38 PRELIMINARY; PRT; 891 AA.  
ID Q93L38  
AC Q93L38  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE MEMBRANE TRANSDUCATOR.  
GN AMT.  
OS Streptomyces griseus.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HHL;  
RX MEDLINE=93209944; PubMed=8458843;  
RA Ueda K., Miyake K., Horinouchi S., Beppu T.;  
RT "A gene cluster involved in aerial mycelium formation in Streptomyces  
griseus encodes proteins similar to the response regulators of two-  
component regulatory systems and membrane translocators.";  
RT J. Bacteriol. 175:2006-2016(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HHL;  
RX MEDLINE=98422460; PubMed=9748440;  
RA Ueda K., Hsieh C.-W., Tosaki T., Shinkawa H., Beppu T., Horinouchi S.;  
RT "Characterization of an A-factor-responsive repressor for amr  
essential for onset of aerial mycelium formation in Streptomyces  
griseus.";  
RL J. Bacteriol. 180:5085-5093(1998).  
DR EMBL: AB006206; BAB62264.1;-  
SQ SEQUENCE 891 AA; 94161 MW; 7CF1487510BCC304 CRC64;

Query Match Best Local Similarity 100.0%; Score 34; DB 2; Length 891;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HCLEH 5  
Db 107 HCLEH 111

RESULT 11  
Q9Y4E4 PRELIMINARY; PRT; 1044 AA.  
ID Q9Y4E4  
AC Q9Y4E4  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE KIA00580 PROTEIN (FRAGMENT).  
GN KIA00580.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-98290545; PubMed-9628581;  
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."  
 RL DNA Res. 5:31-39(1998).  
 DR EMBL: AB011152; BAA25506.1; -  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000198; RhogAP.  
 DR InterPro: IPR001164; Znf\_GCS.  
 DR Pfam: PF01412; ArfGAP; 1.  
 DR Pfam: PF00169; PH; 4.  
 DR Pfam: PF00620; RhogAP; 1.  
 DR PRINTS: PR00405; REVINTFRCTNG.  
 DR SMART: SM00105; ArfGAP; 1.  
 DR SMART: SM00233; PH; 3.  
 DR SMART: SM00324; RhogAP; 1.  
 DR PROSITE: PSS0003; PH\_DOMAIN; 3.  
 FT NON\_TER  
 SQ SEQUENCE 1044 AA; 119464 MW; 68E22A6D6DA53BD CRC64;

Query Match 100.0%; Score 34; DB 4; Length 1044;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
 Db 954 HCLEH 958  
 RESULT 12  
 Q96L70 PRELIMINARY; PRT; 1704 AA.  
 AC Q96L70;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DE AAAP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID-9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Miura K., Jacques K.M., Stauffer S., Kubosaki A., Zhu K., Resau J., Zheng Y., Randazzo P.A.;  
 RT "ARAP1, a point of convergence for phosphoinositide, Arf and Rho signalling, regulates both membrane and actin cytoskeleton remodeling."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY049733; AAL12170.1; -  
 SQ SEQUENCE 1704 AA; 193422 MW; 3770EDE08E42E22C CRC64;

Query Match 100.0%; Score 34; DB 4; Length 1704;  
 Best Local Similarity 100.0%; Pred. No. 16+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
 Db 1614 HCLEH 1618  
 RESULT 13  
 Q924Z9 PRELIMINARY; PRT; 2117 AA.  
 AC Q924Z9;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CDA PEPTIDE SYNTHETASE III (FRAGMENT).  
 GN SCE63.01.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID-1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Saunders D.C., Harris D.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-97000351; PubMed-8843436;  
 RA Redendbach M., Kleser H.M., Denaplatte D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL035707; CAB38876.1; -  
 DR HSSP: P14687; IAMU.  
 DR InterPro: IPR000873; AMP-bind.  
 DR InterPro: IPR001242; DUF4.  
 DR InterPro: IPR00379; Est\_lip\_thioest\_actsite.  
 DR InterPro: IPR003880; Phosphopant\_attach.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00501; AMP-binding; 2.  
 DR Pfam: PF00668; Condensation; 1.  
 DR Pfam: PF00550; PP-binding; 2.  
 DR Pfam: PF00975; Thioesterase; 1.  
 DR PRINTS: PR00154; AMPBINDING.  
 DR PROSITE: PSS0075; ACP\_DOMAIN; 2.  
 DR PROSITE: PSS00455; AMP-BINDING; 2.  
 DR PROSITE: PSS00012; PHOSPHOPANTHETINE; UNKNOWN; 2.  
 FT NON\_TER  
 SQ SEQUENCE 2117 AA; 225115 MW; 30EBF48688DBBEA1 CRC64;

Query Match 100.0%; Score 34; DB 2; Length 2117;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
 Db 466 HCLEH 470  
 RESULT 14  
 Q9PBP3 PRELIMINARY; PRT; 85 AA.  
 AC Q9PBP3;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN XF2097.  
 CN XF2097.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OX NCBI\_TaxID-2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-9A5C;  
 RX MEDLINE-20365717; PubMed-10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Honzel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A., Jr., Nodrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.,  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE004025; AAF64896.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 85 AA; 10186 MW; 361F2727BF73718F CRC64;

Query Match 94.1%; Score 32; DB 16; Length 85;  
 Best Local Similarity 80.0%; Pred. No. 19;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HCLEH 5  
 Db 15 HCMEH 19

RESULT 15  
 Q9JRN7 PRELIMINARY; PRT; 294 AA.  
 AC Q9JRN7;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE HYPOTHETICAL 33.3 KDA PROTEIN.  
 OS Actinobacillus actinomycetemcomitans (Haemophilus  
 OS actinomycetemcomitans).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Actinobacillus.  
 OX NCBI\_TaxID=714;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SUNYAB 75;  
 RA Suzuki N., Nakano Y., Yoshida Y., Nako H., Yamashita Y., Koga T.;  
 RT "Genetic analysis of the gene cluster for the synthesis of serotype a-  
 specific polysaccharide antigen in *A. actinomycetemcomitans*.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB046360; BAB03206.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 294 AA; 33342 MW; ABC3F7E2C8CODEB4 CRC64;

Query Match 94.1%; Score 32; DB 2; Length 294;  
 Best Local Similarity 80.0%; Pred. No. 56;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HCLEH 5  
 Db 133 HCMEH 137

Search completed: August 20, 2002, 11:31:58  
 Job time: 1472 sec



000000

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:07:18 ; Search time 187.18 Seconds  
(without alignments)  
1.780 Million cell updates/sec

Title: US-09-824-286-17  
Perfect score: 19  
Sequence: 1 FNP 3

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_032802.\*  
1: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	3	AAW31654	Human cytokine rec
2	19	100.0	4	AA13261	Cytotoxic Cell Pro
3	19	100.0	5	AA13285	Beta-adrenergic GP
4	19	100.0	6	AA13566	Epitope recognised
5	19	100.0	6	AA15559	Epitope recognised
6	19	100.0	6	AA16944	Rat phospholipase
7	19	100.0	6	AA19813	Transglutaminase C
8	19	100.0	6	AA19833	Human pancreatic p
9	19	100.0	6	AA18407	Amino acid sequenc
10	19	100.0	7	AA15513	Transglutaminase p
11	19	100.0	7	AA16530	T. chalcogramma tr

12	19	100.0	7	AAW52004	Peptide having imm
13	19	100.0	7	AAW51081	Pentillanin cont
14	19	100.0	7	AAW44049	H11 binding site c
15	19	100.0	7	AAW44489	H11 binding site c
16	19	100.0	7	AAW46402	H11 binding site c
17	19	100.0	7	AAW46568	H11 binding site c
18	19	100.0	7	AAW46608	H11 binding site c
19	19	100.0	8	AAW32742	PK40 TAV/neurofila
20	19	100.0	8	AAW72720	MRSA-230 tryptic p
21	19	100.0	8	AAW73401	Human TSH receptor
22	19	100.0	8	AAW73402	Human TSH receptor
23	19	100.0	8	AAW73403	Human TSH receptor
24	19	100.0	8	AAW69539	Human 20K growth h
25	19	100.0	8	AAW27417	PK40 protein kinase
26	19	100.0	9	AAW38111	Hepatitis B virus-
27	19	100.0	9	AAW55144	Transglutaminase p
28	19	100.0	9	AAW55315	Pagrus major trans
29	19	100.0	9	AAW33943	Beta1-adrenergic r
30	19	100.0	9	AAW36467	Cyclitol-ubiquitin
31	19	100.0	9	AAW83244	NPF motif EH domai
32	19	100.0	9	AAW45669	Immunogenic peptid
33	19	100.0	9	AAW45662	Immunogenic peptid
34	19	100.0	9	AAW70721	Hydrophobic sequen
35	19	100.0	9	AAW98143	Transglutaminase c
36	19	100.0	9	AAW3485	Human tyrosine-rel
37	19	100.0	9	AAW3833	Hepatitis C virus
38	19	100.0	9	AAW50026	Mutant antimesothe
39	19	100.0	10	AAW38149	Hepatitis B virus-
40	19	100.0	10	AAW23359	Common receptor su
41	19	100.0	10	AAW69540	Human 20K growth h
42	19	100.0	10	AAW69541	Human 20K growth h
43	19	100.0	10	AAW45717	Immunogenic peptid
44	19	100.0	10	AAW99103	Potato patatin pro
45	19	100.0	10	AAW99104	Potato patatin pro

#### ALIGNMENTS

RESULT 1	
AAW31654	
ID AAW31654 standard; Peptide: 3 AA.	
XX	
AC AAW31654:	
XX	
DT 21-MAY-1998 (first entry)	
XX	
DE Human cytokine receptor gc chain epitope.	
XX	
KW Cytokine receptor; gamma common chain; gc chain; human;	
KW blocking agent; monoclonal antibody; CP.B8; immunological disease;	
KW myasthenia gravis; rheumatoid arthritis; lupus; multiple sclerosis;	
KW insulin-dependent diabetes; inflammatory bowel disease;	
KW sympathetic ophthalmia; uveitis; allergy; asthma; infection;	
KW graft versus host disease; psoriasis; immunosuppressive; therapy;	
KW epitope.	
XX	
OS Homo sapiens.	
XX	
PN WO9743416-A1.	
XX	
PD 20-NOV-1997.	
XX	
PF 09-MAY-1997; 97MO-US07870.	
XX	
PR 10-MAY-1996; 96US-0017466.	
XX	
PA (BIOJ ) BIOGEN INC.	
XX	
PI Benjamin CD, Burkly LC, Hession C, Whitty A;	
XX	
DR WPI; 1998-008885/01.	
XX	

PT Blocking agents of the gamma common chain of cytokine receptors -  
PT particularly monoclonal antibodies, used to induce T cell anergy for  
PT treatment of immunological diseases  
XX  
PS Claim 24; Page 85; 11pp; English.  
XX  
XX This peptide comprises an epitope of the human cytokine receptor  
CC common gamma (gc) chain (see AAW31646) that is recognised by  
CC gc blocking agents of the invention. 5 Such epitopes (see  
CC AAW31650-54) have been identified. The invention provides  
CC compositions and methods for inhibiting cytokine signalling using  
CC gc chain blocking agents for the treatment of immunological  
CC diseases such as myasthenia gravis, rheumatoid arthritis, lupus,  
CC multiple sclerosis, insulin-dependent diabetes, inflammatory bowel  
CC disease, symptomatic ophthalmia, uveitis, allergy, asthma,  
CC parasitic infection, graft vs. host disease or psoriasis. A  
CC preferred gc blocking agent is Mab CP.88 or its Fab fragment (see  
CC also AAW31647-48).  
XX  
SQ Sequence 3 AA;  
  
Query Match 100.0%; Score 19; DB 19; Length 3;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FNP 3  
   | | |  
DB 1 fnp 3  
  
RESULT 2  
AA13261  
ID AA13261 standard; Protein; 4 AA.  
XX  
XX AA13261;  
XX  
XX 11-OCT-1991 (first entry)  
XX  
XX Cytotoxic Cell Protease Inhibitor EP2372.  
XX  
XX hCCP inhibitor; cytotoxic T-lymphocytes; CTL.  
XX  
XX Homo sapiens.  
XX  
XX MO9110685-A.  
XX  
XX 25-JUL-1991.  
XX  
XX 17-JAN-1991; 91MO-US00340.  
XX  
XX 19-JAN-1990; 90US-0467880.  
XX  
XX (SERA-) SERAGEN INC.  
XX  
XX Bleackley RC, Lobe CG, Paetkau VH, James MN, Murphy M;  
PI WPI: 1991-237989/32.  
XX  
XX DNA vectors, and inhibitors of cytotoxic cell protease - for  
PT treatment of auto-immune diseases e.g. pernicious anaemia,  
PT rheumatoid arthritis, allo-graft rejection etc.  
XX  
XX Claim 9; Fig 11; 62pp; English.  
XX  
XX cDNAs encoding murine and human CCP's were isolated and sequenced.  
CC The amino acid sequences they encode were deduced and computer  
CC analysis of the protein structure was performed. This protease  
CC inhibitor is one of nine tetrapeptides of the invention which act  
CC as competitive inhibitors. They are based on the protease cleavage  
CC sites of CCP substrate molecules which were inferred from the  
CC analysis of the protease binding pocket.  
CC See AAQ12862-6 and AA13254-R13262.

XX  
SQ Sequence 4 AA;  
  
Query Match 100.0%; Score 19; DB 12; Length 4;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FNP 3  
   | | |  
DB 1 fnp 3  
  
RESULT 3  
AA13295  
ID AA13295 standard; peptide; 5 AA.  
XX  
XX AA13295;  
XX  
XX 28-SEP-1999 (first entry)  
XX  
XX Beta-adrenergic GPCR consensus peptide motif.  
XX  
XX Gene characterisation; recombinase; homology clamp; homology motif tag;  
XX HMT; genetic manipulation; genetic engineering; gene transcription; GPCR;  
XX drug target; homologous recombination; target identification; TGF-beta;  
XX G-protein coupled receptor; transforming growth factor-beta.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 1  
FT /Label= Ala or Gly  
FT Misc-difference 5  
FT /Label= Ile or Leu  
XX  
XX WO9937755-A2.  
XX  
XX 29-JUL-1999.  
XX  
XX 11-DEC-1998; 98MO-US26498.  
XX  
XX 11-DEC-1997; 97US-0070734.  
XX  
XX (PANG-) PANGENE CORP.  
XX  
XX Lehman CW, Patl S, Zarling D, Zeng H;  
PI WPI: 1999-458689/38.  
XX  
XX New compositions and methods for targeting sequence modifications in  
PT related family genes  
XX  
XX Disclosure: Fig 1B; 46pp; English.  
XX  
XX The invention provides compositions and methods for the evaluation and  
CC characterisation of individual and sets of genes in disease states. The  
CC composition comprises at least one recombinase and at least two single-  
CC stranded targeting polynucleotides which are substantially complementary  
CC to each other and each having a consensus homology clamp for a gene  
CC family i.e. a homology motif tag (HMT). The composition is useful in kit  
CC form which include the composition as libraries or pools of degenerate  
CC cssDNA probes along with other reagents such as recombinase etc. The  
CC methods and compositions are used for inactivation of a gene family gene  
CC i.e. exogenous targeting polynucleotides can be used to inactivate,  
CC decrease or alter the biological activity of one or more genes in a cell  
CC (or transgenic nonhuman animal or plant). This is useful for generating  
CC animal models of disease, or in the elucidation of gene function and  
CC activity. Alternatively, the biological activity of the wild-type gene  
CC may be either decreased or the wild-type activity altered to mimic  
CC disease states. This includes genetic manipulation of non-coding gene  
CC sequences that affect the transcription of genes, including promoter,  
CC repressors, enhancers and transcriptional activating sequences. The

CC compositions are useful in identifying new members of gene families  
 CC which may be useful in functional genomic studies as well as in  
 CC identification of new drug targets. HMTs used in homologous recombination  
 CC methods can generate animals that have a wide variety of mutations in a  
 CC wide variety of related genes, potentially resulting in a wide variety of  
 CC phenotypes including those related to disease states. This may also be  
 CC done on a cellular level to identify genes involved in cellular  
 CC phenotypes i.e. target identification. Sequences AAY31248-308 represent  
 CC consensus peptide motifs of the G-protein coupled receptor (GPCR) gene  
 CC family and transforming growth factor-beta (TGF-beta) gene family.  
 XX  
 SQ Sequence 5 AA;

Query Match 100.0%; Score 19; DB 20; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
 |||  
 Db 2 fnp 4

RESULT 4  
 AAR55566  
 ID AAR55566 standard; Protein; 6 AA.  
 XX

AC AAR55566;

DT 10-FEB-1995 (first entry)

DE Epitope recognised by MAB 87-55/02/2, isolated from random library.

KM epitope; binding specificity; random hexapeptide; ligand screening;  
 KW outer membrane protein; ompa; cytomagalovirus; CMV; ppl50 protein.  
 XX

OS Synthetic.

PN EP603672-A.

PD 29-JUN-1994.

PF 10-DEC-1993; 93EP-0119921.

PR 23-DEC-1992; 92DE-4243770.

PA (BEHW ) BEHRINGWERKE AG.

PI Grundmann U, Wessel T, Zetlmeissl G;

DR WPI; 1994-201750/25.

XX  
 PT Prepn. of peptide(s) with specific binding properties - by  
 PT expressing oligo-nucleotide(s) in bacteria or animal cells, as  
 PT surface proteins, then screening with specific receptor attached  
 PT to solid phase  
 XX

PS Example 4; Page 10; 25pp; German.

XX A library of random hexapeptides was screened for epitopes which  
 CC are recognised by the Monoclonal antibody MAB 87-55/02/2. The MAB  
 CC is known to recognise an epitope of the ppl50 protein of  
 CC cytomagalovirus. Antibody screening identified 5 hexapeptide motifs  
 CC (AAR55566-R55570), all of which differed from the wild-type epitope.  
 XX  
 SQ Sequence 6 AA;

Query Match 100.0%; Score 19; DB 15; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3

Db 1 fnp 3

RESULT 5  
 AAR55569  
 ID AAR55569 standard; Protein; 6 AA.  
 XX

AC AAR55569;

DT 10-FEB-1995 (first entry)

DE Epitope recognised by MAB 87-55/02/2, isolated from random library.

KM epitope; binding specificity; random hexapeptide; ligand screening;  
 KW outer membrane protein; ompa; cytomagalovirus; CMV; ppl50 protein.  
 XX

OS Synthetic.

PN EP603672-A.

PD 29-JUN-1994.

PF 10-DEC-1993; 93EP-0119921.

PR 23-DEC-1992; 92DE-4243770.

PA (BEHW ) BEHRINGWERKE AG.

PI Grundmann U, Wessel T, Zetlmeissl G;

DR WPI; 1994-201750/25.

XX  
 PT Prepn. of peptide(s) with specific binding properties - by  
 PT expressing oligo-nucleotide(s) in bacteria or animal cells, as  
 PT surface proteins, then screening with specific receptor attached  
 PT to solid phase  
 XX

PS Example 4; Page 10; 25pp; German.

XX A library of random hexapeptides was screened for epitopes which  
 CC are recognised by the Monoclonal antibody MAB 87-55/02/2. The MAB  
 CC is known to recognise an epitope of the ppl50 protein of  
 CC cytomagalovirus. Antibody screening identified 5 hexapeptide motifs  
 CC (AAR55566-R55570), all of which differed from the wild-type epitope.  
 XX

SQ Sequence 6 AA;

Query Match 100.0%; Score 19; DB 15; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
 |||  
 Db 1 fnp 3

RESULT 6  
 AAM69443  
 ID AAM69443 standard; Protein; 6 AA.  
 XX

AC AAM69443;

DT 10-DEC-1998 (first entry)

DE Rat phospholipase A1 fragment.

KW Phospholipase A1; phosphatidyl-serine hydrolysis; serine phospholipid;  
 KW PLAI; enzyme; rat.  
 XX

OS Rattus sp.

XX

PN JP10201479-A.  
XX  
XX 04-AUG-1998.  
XX  
XX 23-JAN-1997; 97JP-0024269.  
XX  
XX 23-JAN-1997; 97JP-0024269.  
XX  
XX (TORA ) TORAY IND INC.  
XX  
XX WPI: 1998-474493/41.  
XX  
XX Phospholipase A1 and nucleic acid coding for it - having substrate  
PT specifically against serine phospholipid  
XX  
XX Example 1; Page 5; 13pp; Japanese.  
XX  
XX This sequence is a fragment of the rat phospholipase A1 (PLA1) of the  
CC invention. PLA1 is capable of hydrolysing phosphatidyl-serine, and has  
CC substrate specificity for serine phospholipids.  
XX  
XX Sequence 6 AA;  
SQ

Query Match 100.0%; Score 19; DB 19; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
|||  
Db 1 fnp 3

RESULT 7  
AAW98139  
ID AAW98139 standard; Peptide: 6 AA.  
XX  
XX AAW98139;  
AC  
XX 05-JUL-1999 (first entry)  
DT  
XX Transglutaminase conserved peptide.  
DE  
XX Transglutaminase: diagnosis; therapy.  
KW  
XX Homo sapiens.  
OS  
XX Mus sp.  
OS  
XX Rattus sp.  
XX  
XX WO9910507-A1.  
PN  
XX 04-MAR-1999.  
PD  
XX 28-AUG-1998; 98WO-US17857.  
XX  
XX 29-AUG-1997; 97US-0920919.  
PE  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
PA  
XX Aeschlimann DP, Mosher DF;  
PI  
XX WPI: 1999-190622/16.  
DE  
XX  
XX Novel transglutaminase and new degenerate and specific primers -  
PT useful in diagnostic assays for amplification and characterization  
PT of transglutaminase genes and proteins  
XX  
XX Disclosure; Page 14; 68pp; English.  
XX  
XX The present sequence is a conserved region of human, mouse and  
CC rat transglutaminases. Short (18-mer) oligonucleotides based  
CC on this peptide were unsuccessful as PCR primers for amplifying  
CC transglutaminase gene sequences. The invention provides

CC degenerate and specific primers (see AAX24958, AAX24960-73) for  
CC the PCR amplification of known and novel transglutaminase  
CC genes, including the novel human transglutaminase Tgx gene (see  
CC AAX24956).  
XX  
XX  
SQ Sequence 6 AA;

Query Match 100.0%; Score 19; DB 20; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
|||  
Db 2 fnp 4

RESULT 8  
AAV95393  
ID AAV95393 standard; Peptide: 6 AA.  
XX  
XX AAV95393;  
AC  
XX 25-SEP-2000 (first entry)  
DT  
XX Human pancreatic polypeptide zslg66 motif 2.  
DE  
XX zslg66; human; pancreas; infection; diabetes; diagnosis; therapy.  
KW  
XX Homo sapiens.  
OS  
XX WO200036104-A1.  
PN  
XX 22-JUN-2000.  
PD  
XX 14-DEC-1999; 99WO-US29669.  
PF  
XX 16-DEC-1998; 98US-0212947.  
PR  
XX (ZYMO ) ZYMOGENETICS INC.  
PA  
XX Sheppard PO;  
PI  
XX WPI: 2000-475395/41.  
DR

XX Nucleic acids encoding pancreatic zslg66 proteins useful for  
XX identifying compounds that may be used to treat, for example microbial  
PT infections -  
PT  
XX  
XX Disclosure; Page 14; 108pp; English.  
PS

XX The present sequence is that of peptide motif 2 of zslg66 (see  
CC AAV95391), a novel human secreted pancreatic polypeptide. The motif  
CC can be used to design oligonucleotide primers (see AAX9885) useful  
CC for identifying related sequences. zslg66 polypeptides can be  
CC used in the prevention, treatment and diagnosis of diseases  
CC associated with inappropriate zslg66 expression, to raise  
CC antibodies, and to screen for modulators (agonists and antagonists)  
CC of zslg66 activity.  
CC  
XX

SQ Sequence 6 AA;

Query Match 100.0%; Score 19; DB 21; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
|||  
Db 3 fnp 5

RESULT 9

AAy84078  
ID AAY84078 standard; Peptide; 6 AA.  
XX  
AC AAY84078;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Amino acid sequence of the C-terminal of a defensin fusion protein.  
XX  
KW Antimicrobial protein; AMP1; transgenic plant; linker propeptide;  
KM protein expression; plant defensin; RsaFP2; antifungal protein; AFP2.  
XX  
OS Synthetic.  
XX  
PN WO200011175-A1.  
XX  
PD 02-MAR-2000.  
XX  
PF 17-AUG-1999; 99WO-GB02716.  
XX  
PR 18-AUG-1998; 98GB-0018001.  
XX 04-DEC-1998; 98GB-0026753.  
XX  
PA (ZENE ) ZENECA LTD.  
XX  
PI Broekaert WF, Francois IEJA, De Bolle MFC, Evans IJ, Ray JA;  
XX WPI; 2000-246564/21.  
XX  
XX  
XX Improving expression of polypeptides in plants involves coexpression of  
PT two or more proteins in plants within a single transcription unit -  
XX  
XX  
PS Example 10; Page 40; 151pp; English.  
XX  
CC The present sequence represents the C-terminal of a fusion protein  
CC of the invention, comprising the Dahlia antimicrobial protein (AMP) 1  
CC and an antifungal protein (RsaFP2), linked by a linker propeptide. The  
CC specification describes methods for improving expression levels of one  
CC or more proteins in a transgenic plant. The method comprises inserting a  
CC DNA sequence having a promoter region operably linked to two or more  
CC protein encoding regions separated by a DNA sequence coding for a  
CC linker propeptide and a terminator region. The method is used to  
CC produce proteins in plants. The linker propeptide comprising a cleavage  
CC site, whereby the expressed polypeptide is post-translationally  
CC processed into the component protein molecules. The propeptide sequence  
CC is rich in amino acids A, V, S and T and contains dipeptidic sequences  
CC consisting of either two acidic, two basic or one acidic and one basic  
CC residue as a cleavable linker sequence.  
XX  
SQ Sequence 6 AA:  
  
Query Match 100.0%; Score 19; DB 21; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Oy 1 FNP 3  
|||  
Db 3 fnp 5  
  
RESULT 10  
AAR5138  
ID AAR5138 standard; peptide; 7 AA.  
XX  
AC AAR5138;  
XX  
DT 16-DEC-1994 (first entry)  
XX  
DE Transglutaminase peptide.  
XX  
KW Transglutaminase; TGase; fish; liver; Teragra chalcogramma;  
KM Pagrus major; enzyme; acyl transfer; gamma-carboxamide;  
XX

KM glutamine.  
XX  
OS Teragra chalcogramma.  
XX  
PN JP06113844-A.  
XX  
PD 26-APR-1994.  
XX  
PF 15-DEC-1992; 92JP-0334224.  
XX  
PR 19-AUG-1992; 92JP-0220296.  
XX  
PA (AJIN ) AJINOMOTO KK.  
XX  
DR WPI; 1994-172742/21.  
XX  
XX  
XX Compn. contg. trans-glutaminase derived from fish liver - useful  
PT for prepn. of protein-gelled prod., esp. food prod.  
XX  
PS Disclosure; Page 10; 26pp; Japanese.  
XX  
CC A compsn. contains at least 0.5 units per mg protein of a novel  
CC transglutaminase (TGase). The TGase is derived from fish liver  
CC (esp. Teragra chalcogramma and Pagrus major) and catalyses the  
CC acyl transfer reaction between the gamma-carboxamide gp. of the  
CC glutamine residue of a peptide chain and various prim. amines.  
XX  
XX TGase is useful for prepn. of protein-gelled prods., esp. food.  
XX  
SQ Sequence 7 AA:  
  
Query Match 100.0%; Score 19; DB 15; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Oy 1 FNP 3  
|||  
Db 5 fnp 7  
  
RESULT 11  
AAR65309  
ID AAR65309 standard; Peptide; 7 AA.  
XX  
AC AAR65309;  
XX  
DT 31-MAY-1996 (first entry)  
XX  
DE T. chalcogramma transglutaminase peptide fragment.  
XX  
KW transglutaminase; fish; recombinant production; Pagrus major;  
KM Theragra chalcogramma; Parailichthys olivaceus; Oncorhynchus keta.  
XX  
PN JP07023787-A.  
XX  
PD 27-JAN-1995.  
XX  
PF 13-JUL-1993; 93JP-0172998.  
XX  
PR 13-JUL-1993; 93JP-0172998.  
XX  
PA (AJIN ) AJINOMOTO KK.  
XX  
DR WPI; 1995-100948/14.  
XX  
XX  
XX A fish-originated trans-glutaminase gene - can be produced in large  
PT amounts at low economic cost.  
XX  
PS Example 12; Page 42; 94pp; Japanese.  
XX  
CC AAR65307-12 are fragments of a Theragra chalcogramma protein which has

CC transglutaminase (TG) activity. The TG's of the invention can be  
 CC isolated from *Pagrus major*, *Theragra chalcogramma*, *Paralichthys olivaceus*  
 CC or *Oncorhynchus keta*. Fish-originated TG's are economical to produce in  
 CC large quantities.  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 19; DB 16; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 III  
 DB 5 fnp 7

RESULT 12  
 AAM52004  
 ID AAM52004 standard; peptide: 7 AA.

AC AAM52004;  
 XX  
 DT 21-SEP-1998 (first entry)  
 XX

DE Peptide having immunomodulating activity.

KW Immunomodulating activity; cancer; autoimmune disorder; infection;  
 KW cysteine analogue.

OS Synthetic.

FT Key Location/Qualifiers  
 FT Modified-site 6 /note= "pencillamine"

PN W09812219-A1.

PD 26-MAR-1998.

PF 15-SEP-1997; 97WO-SE01554.

PR 23-SEP-1996; 96SE-0003468.

PA (ASTR ) ASTRA AB.

PI Bergstrand H, Eriksson T, Lindvall M, Saernstrand B;

DR WPI; 1998-271701/24.

PT New peptide(s) containing 3-30 amino acids - are immunomodulatory  
 PT compounds useful in treatment of cancer or autoimmune disorders  
 XX  
 PS Example 138; Page 54; 75pp; English.

CC The invention relates to methionine, penicillamine and cysteine-analogue  
 CC containing peptides which are immunomodulatory (either immuno-  
 CC inhibitory or immunostimulatory) compounds. They may be used in treatment  
 CC of cancers, infections associated with autoimmune disease, autoimmune  
 CC diseases, asthma, rhinitis, fibrosis, chronic bronchitis, hepatitis,  
 CC post-infectious anergy, AIDS, HIV or post-traumatic immunological  
 CC anergy. The peptides may also be used as adjuvants in vaccines, or in  
 CC formulations designed to prevent rejection of transplanted organs. The  
 CC peptides are absorbable by the epithelial cell lining. Thus the dosage  
 CC required to achieve the therapeutic effect by oral, nasal or  
 CC intratracheal delivery can be much lower than that required to produce  
 CC the same effect via, e.g., parenteral administration. The present  
 CC sequence represents a peptide having immunomodulating activity.  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 19; DB 19; Length 7;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 III  
 DB 3 fnp 5

RESULT 13  
 AAM51081  
 ID AAM51081 standard; peptide: 7 AA.

AC AAM51081;  
 XX  
 DT 05-AUG-1998 (first entry)  
 XX

DE Peicillamine containing peptide #44 having immuno-modulatory effect.

KW Pencillamine; immuno-modulatory; epithelial cell lining; cancer;  
 KW auto-immune disease.

OS Synthetic.

FT Key Location/Qualifiers  
 FT Modified-site 6 /note= "penicillamine"

PN W09812215-A1.

PD 26-MAR-1998.

PF 15-SEP-1997; 97WO-SE01549.

PR 23-SEP-1996; 96SE-0003462.

PA (ASTR ) ASTRA AB.

PI Bergstrand H, Eriksson T, Lindvall M, Saernstrand B;

DR WPI; 1998-230309/20.

PT Peptide(s) which contain penicillamine residue - are  
 PT immuno-modulatory compounds useful in treating, e.g. cancer or  
 PT auto-immune disorders  
 XX  
 PS Example 42; Page 31; 71pp; English.

CC The invention relates to penicillamine containing peptides having an  
 CC immuno-modulatory (either immuno-inhibitory or immuno-stimulatory)  
 CC effect. The peptides may be used in treatment of cancers, infections  
 CC associated with autoimmune disease, autoimmune diseases (e.g. non-obese  
 CC diabetes, systemic lupus erythematosus, scleroderma, Sjogren's syndrome,  
 CC dermatomyositis, multiple sclerosis, rheumatoid arthritis, arterio-  
 CC sclerosis or psoriasis), asthma, rhinitis, fibrosis, chronic bronchitis,  
 CC hepatitis, post-infectious anergy, AIDS, HIV or post-traumatic  
 CC immunological anergy. The peptides may also be used as adjuvants in  
 CC vaccines, or in formulations designed to prevent rejection of  
 CC transplanted organs. The peptides are absorbable by the epithelial cell  
 CC lining. Thus the dosage required to achieve the therapeutic effect by  
 CC oral, nasal or intratracheal delivery can be significantly lower than  
 CC that required to produce the same effect via, e.g. parenteral  
 CC administration. The present sequence represents a penicillamine  
 CC containing peptide.  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 19; DB 19; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 III



Db 3 fnp 5

## RESULT 14

ID AAM44049 standard; Peptide; 7 AA.

AC AAM44049;

DT 25-OCT-2001 (first entry)

DE H11 binding site consensus conforming peptide (CCP) #320.

KW Antigen-binding; tumour; diagnosis; stress protein-peptide complex; SPPC;

KW immunogenically cross-reactive; cancer; immunogenic cancer cell;

KW cytostatic; vaccine; tumour-specific immunogenic response inducer;

KW astrocytoma; fibrosarcoma; myxosarcoma; liposarcoma; oligodendroglioma;

KW ependymoma; medulloblastoma; primitive neural ectodermal tumour.

XX Homo sapiens.

OS Synthetic.

PN CA2290722-A1.

PD 08-JUN-2001.

PF 08-DEC-1999; 99CA-2290722.

PR 08-DEC-1999; 99CA-2290722.

PA (NOVO-) NOVOPHARM BIOTECH INC.

XX Kaplan HA, Malti PK, Fast DG, Herman W, Dan MD, Lewis KE;

PI Entwistle JM, MacDonald GC;

DR WPI; 2001-425937/46.

XX Composition useful for treating and diagnosing cancer, comprises stress

PT protein-peptide complexes associated with tumor, and isolated

PT antigen-binding fragments of an antibody that binds specifically to the

XX complex

XX Example 4; Page 102; 154pp; English.

PS The present invention describes a composition (I) comprising stress

CC protein-peptide complexes (SPPC) associated with tumors that is

CC specifically immunogenically cross-reactive with cell surface-associated

CC SPPCs specific to target cancer (TC). Also described is an isolated

CC antigen-binding fragment of an antibody that binds specifically to SPPCs

CC or a population of different SPPCs consisting of immunogenic cancer cell

CC surface-associated SPPC of TC. (I) has cytostatic activity and can be

CC used in vaccine production and as a tumour-specific immunogenic response

CC inducer. (I) is useful for treating 71 types of cancers or tumours in a

CC subject, such as astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma,

CC oligodendroglioma, ependymoma, medulloblastoma, and primitive neural

CC ectodermal tumour (PNET). (I) is useful as cancer immunogen including

CC vaccines. (I) is useful for diagnostic and palliative use, for detecting

CC or imaging cancer cells, and to monitor the course of amelioration of

CC malignancy in an individual. AAM43707 to AAM47109 represent peptides

CC which are used in the exemplification of the present invention.

XX Sequence 7 AA;

SQ

Query Match 100.0%; Score 19; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
111  
Db 4 fnp 6

## RESULT 15

ID AAM44489 standard; Peptide; 7 AA.

AC AAM44489;

DT 25-OCT-2001 (first entry)

DE H11 binding site consensus conforming peptide (CCP) #760.

KW Antigen-binding; tumour; diagnosis; stress protein-peptide complex; SPPC;

KW immunogenically cross-reactive; cancer; immunogenic cancer cell;

KW cytostatic; vaccine; tumour-specific immunogenic response inducer;

KW astrocytoma; fibrosarcoma; myxosarcoma; liposarcoma; oligodendroglioma;

KW ependymoma; medulloblastoma; primitive neural ectodermal tumour.

XX Homo sapiens.

OS Synthetic.

PN CA2290722-A1.

PD 08-JUN-2001.

PF 08-DEC-1999; 99CA-2290722.

PR 08-DEC-1999; 99CA-2290722.

PA (NOVO-) NOVOPHARM BIOTECH INC.

XX Kaplan HA, Malti PK, Fast DG, Herman W, Dan MD, Lewis KE;

PI Entwistle JM, MacDonald GC;

DR WPI; 2001-425937/46.

XX Composition useful for treating and diagnosing cancer, comprises stress

PT protein-peptide complexes associated with tumor, and isolated

PT antigen-binding fragments of an antibody that binds specifically to the

XX complex

XX Example 4; Page 103; 154pp; English.

PS The present invention describes a composition (I) comprising stress

CC protein-peptide complexes (SPPC) associated with tumors that is

CC specifically immunogenically cross-reactive with cell surface-associated

CC SPPCs specific to target cancer (TC). Also described is an isolated

CC antigen-binding fragment of an antibody that binds specifically to SPPCs

CC or a population of different SPPCs consisting of immunogenic cancer cell

CC surface-associated SPPC of TC. (I) has cytostatic activity and can be

CC used in vaccine production and as a tumour-specific immunogenic response

CC inducer. (I) is useful for treating 71 types of cancers or tumours in a

CC subject, such as astrocytoma, fibrosarcoma, myxosarcoma, liposarcoma,

CC oligodendroglioma, ependymoma, medulloblastoma, and primitive neural

CC ectodermal tumour (PNET). (I) is useful as cancer immunogen including

CC vaccines. (I) is useful for diagnostic and palliative use, for detecting

CC malignancy in an individual. AAM43707 to AAM47109 represent peptides

CC which are used in the exemplification of the present invention.

XX Sequence 7 AA;

SQ

Query Match 100.0%; Score 19; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
111  
Db 4 fnp 6

Search completed: August 20, 2002, 11:07.19  
Job time: 6104 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:10:39 ; Search time 100.84 Seconds  
(without alignments)  
2.859 Million cell updates/sec

Title: US-09-824-286-17

Perfect score: 19

Sequence: 1 FNP 3

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	19	100.0	15	2 S08209	hypothetical prote
2	19	100.0	15	2 A61612	allatostatin - tob
3	19	100.0	20	2 A54519	tubulin alpha chain
4	19	100.0	24	2 I45851	adenosylmethionine
5	19	100.0	25	2 T46841	hypothetical prote
6	19	100.0	25	2 S69139	bumetanide-binding
7	19	100.0	26	2 PS0111	beta-galactoside b
8	19	100.0	27	2 H44616	homeotic protein H
9	19	100.0	27	2 I44616	homeotic protein H
10	19	100.0	29	2 A55445	repy protein - Esc
11	19	100.0	31	2 G82278	hypothetical prote
12	19	100.0	33	2 S65599	hevein - Para rubb
13	19	100.0	35	2 S42876	probable succinate
14	19	100.0	35	2 H82103	hypothetical prote
15	19	100.0	35	2 A55850	7alpha-hydroxyster
16	19	100.0	37	2 B45187	homeotic protein G
17	19	100.0	37	2 D95008	hypothetical prote
18	19	100.0	40	2 S77793	transketolase (EC
19	19	100.0	40	2 B24095	gamma-hordein 3 -
20	19	100.0	42	2 T06840	photosystem I chain
21	19	100.0	42	2 A05049	hypothetical prote
22	19	100.0	45	2 C82233	hypothetical prote
23	19	100.0	45	2 E49050	T-cell surface gly
24	19	100.0	46	2 S12084	hypothetical prote
25	19	100.0	48	2 P00617	DNA-directed DNA p
26	19	100.0	49	2 S69087	annexin VI, mitoch
27	19	100.0	49	2 A61063	hypothetical secre
28	19	100.0	50	2 P90761	hypothetical prote
29	19	100.0	51	1 A61125	insulin - American

30	19	100.0	51	1 R6C46	ribosomal protein
31	19	100.0	51	2 C46353	ORF3 protein - coc
32	19	100.0	52	2 S06492	beta-galactoside-b
33	19	100.0	52	2 S63981	immune-inducible p
34	19	100.0	53	2 S66603	xanthine oxidase (
35	19	100.0	53	2 S53698	ribosomal protein
36	19	100.0	53	2 B55392	abacsin precursor
37	19	100.0	53	2 A96216	hypothetical prote
38	19	100.0	54	2 A72206	hypothetical prote
39	19	100.0	55	2 E90626	ATP synthase F0 ch
40	19	100.0	55	2 D37334	D-mandelate dehydr
41	19	100.0	55	2 AD2382	hypothetical prote
42	19	100.0	55	2 G98149	hypothetical prote
43	19	100.0	56	2 P50424	H11 protein - Toxo
44	19	100.0	57	2 T03092	hypothetical prote
45	19	100.0	57	2 T22546	hypothetical prote

#### ALIGNMENTS

RESULT 1  
S08209  
hypothetical protein 2 - garden pea  
N:Alternate names: phytochrome  
C:Species: Pisum sativum (garden pea)  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Sep-1997  
C:Accession: S08209  
R:Sato, N.  
Plant Mol. Biol. 11, 697-710, 1988  
A:Title: Nucleotide sequence and expression of the phytochrome gene in Pisum sativum  
A:Reference number: S08856  
A:Accession: S08209  
A:Molecule type: DNA  
A:Residues: 1-15 <SAT>  
A:Cross-references: EMBL:X14077; NID:q20836; PID:q20838  
C:Genetics:  
A:Gene: phy

Query Match 100.0%; Score 19; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
DB 11 FNP 13  
RESULT 2  
A61612  
allatostatin - tobacco hornworm  
C:Species: Manduca sexta (tobacco hornworm)  
C:Date: 21-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 11-Jul-1997  
C:Accession: A61612  
R:Kramer, S.J.; Toschi, A.; Miller, C.A.; Kataoka, H.; Quistad, G.B.; Li, J.P.; Carn  
Proc. Natl. Acad. Sci. U.S.A. 88, 9458-9462, 1991  
A:Title: Identification of an allatostatin from the tobacco hornworm Manduca sexta.  
A:Reference number: A61612; MUID:92052112  
A:Accession: A61612  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-15 <KRA>  
C:Keywords: neuropeptide; pyroglutamic acid  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 100.0%; Score 19; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
DB 11

Db 9 FNP 11

## RESULT 3

A:Accession: A54519

C:Species: Leishmania enriettii (fragment)

C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 13-Aug-1999

C:Accession: A54519

C:Title: Isolation and characterization of an alpha-tubulin gene from Leishmania enriettii

A:Reference number: A54519; MUID:84142075

A:Accession: A54519

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-20 &lt;MIR&gt;

A:Cross-references: GB:M28001; NID:g159409; PIDN:AAA9273.1; PID:g159410

C:Superfamily: tubulin

Query Match

Best Local Similarity 100.0%; Score 19; DB 2; Length 20;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FNP 3

Db 17 FNP 19

## RESULT 4

A:Accession: I45851

C:Species: Bos primigenius taurus (cattle)

C:Date: 15-Oct-1996 #sequence\_revision 15-Oct-1996 #text\_change 22-Jun-1999

C:Accession: I45851

C:Title: Isolation of a cDNA clone encoding S-adenosylmethionine decarboxylase: expressed

A:Reference number: I45851; MUID:86304300

A:Accession: I45851

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-24 &lt;MAC&gt;

A:Cross-references: GB:M14289; NID:g162624; PIDN:AAA30360.1; PID:g552313

C:Superfamily: eukaryotic adenosylmethionine decarboxylase

C:Keywords: carbon-carbon lyase; carboxy-lyase; polyamine biosynthesis

Query Match

Best Local Similarity 100.0%; Score 19; DB 2; Length 24;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FNP 3

Db 15 FNP 17

## RESULT 5

A:Accession: T46841

C:Species: Clostridium acetobutylicum

C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000

C:Accession: T46841

C:Title: The kdp system of Clostridium acetobutylicum: cloning, sequencing, and transcri

A:Reference number: 224106; MUID:97369818

A:Accession: T46841

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-25 &lt;TRD&gt;

A:Cross-references: EMBL:U44892; NID:g2275246; PIDN:AAC45476.1; PID:g2275248

A:Experimental source: DSM 79

## Query Match

Best Local Similarity 100.0%; Score 19; DB 2; Length 25;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FNP 3

Db 20 FNP 22

## RESULT 6

A:Accession: S69139

C:Species: Rattus norvegicus (Norway rat)

C:Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 17-Mar-1999

C:Accession: S69139

C:Title: The 60-kDa bumetanide-binding protein from rat liver membranes is a catalase

A:Reference number: S69139; MUID:95220382

A:Accession: S69139

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-25 &lt;OTY&gt;

Query Match

Best Local Similarity 100.0%; Score 19; DB 2; Length 25;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FNP 3

Db 19 FNP 21

## RESULT 7

A:Accession: PS0111

C:Species: Callithrix jacchus (common marmoset)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 07-May-1999

C:Accession: PS0111

C:Title: Purification and properties of a beta-galactoside-binding lectin from neonat

A:Reference number: PS0111; MUID:90256719

A:Accession: PS0111

A:Molecule type: protein

A:Residues: 1-26 &lt;OHS&gt;

Query Match

Best Local Similarity 100.0%; Score 19; DB 2; Length 26;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FNP 3

Db 24 FNP 26

## RESULT 8

A:Accession: H44616

C:Species: Petromyzon marinus (sea lamprey)

C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 15-Oct-1999

C:Accession: H44616

C:Title: Expansion of the Hox gene family and the evolution of chordates.

A:Reference number: A44616; MUID:93317669

A:Accession: H44616

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-27 <PEN>  
A:Cross-references: GB:U14901; NID:g290821; PIDN:AAA02536.1; PID:g290822  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

Query Match 100.0%; Score 19; DB 2; Length 27;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
|||  
Db 2 FNP 4

RESULT 9  
144616  
homeotic protein Hox 8 (clone 8r) - sea lamprey (fragment)  
C:Species: Petromyzon marinus (sea lamprey)  
C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 15-Oct-1999  
C:Accession: U144616  
R:Pendleton, J.W.; Nagai, B.K.; Murtha, M.T.; Ruddle, F.H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6300-6304, 1993  
A:Title: Expansion of the Hox gene family and the evolution of chordates.  
A:Reference number: A44616; MUID:93317669  
A:Accession: U144616  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-27 <PEN>  
A:Cross-references: GB:U14903; NID:g290823; PIDN:AAA02538.1; PID:g290824  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

Query Match 100.0%; Score 19; DB 2; Length 27;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
|||  
Db 2 FNP 4

RESULT 10  
A35445  
rep1 protein - Escherichia coli plasmids  
N:Alternate names: rep1 protein  
C:Species: Escherichia coli  
C:Date: 14-Sep-1990 #sequence\_revision 14-Sep-1990 #text\_change 19-Oct-1995  
C:Accession: A35445; MUID:90285201  
R:Hama, C.; Takizawa, T.; Moriwaki, H.; Mizobuchi, K.  
J. Biol. Chem. 265, 10666-10673, 1990  
A:Title: Role of leader peptide synthesis in rep2 gene expression of the Col1b-P9 plasmid  
A:Reference number: A35445; MUID:90285201  
A:Accession: A35445  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-29 <HAM>  
A:Experimental source: plasmid Col1b-P9  
J. Paszkiewicz, J.; Wilson, I.W.; Piltard, A.J.  
R. Bacteriol. 174, 2376-2383, 1992  
A:Title: Mutations affecting translational coupling between the rep genes of an IncB plasmid  
A:Reference number: A42382; MUID:92202169  
A:Accession: A42382  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-29 <PR>  
A:Cross-references: GB:M93063; GB:M38522  
A:Experimental source: plasmid PM0720  
C:Genetics:  
A:Genome: plasmid

Query Match 100.0%; Score 19; DB 2; Length 29;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
|||  
Db 7 FNP 9

RESULT 11  
G82278  
hypothetical protein VC0805 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: G82278  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers,  
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: G82278  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-31 <HEI>  
A:Cross-references: GB:AE004166; GB:AE003852; NID:g96555259; PIDN:AAF93969.1; GSPDB:GT  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC0805  
A:Map position: 1

Query Match 100.0%; Score 19; DB 2; Length 31;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
|||  
Db 17 FNP 19

RESULT 12  
S65599  
hevein - Para rubber tree (fragments)  
C:Species: Hevea brasiliensis (Para rubber tree)  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 07-May-1999  
C:Accession: S65599  
R:Soedjanatmadja, U.M.S.; Subroto, T.; Belintema, J.J.  
FEBS Lett. 363, 211-213, 1995  
A:Title: Processed products of the hevein precursor in the latex of the rubber tree (Hevea brasiliensis)  
A:Reference number: S65599; MUID:95253547  
A:Accession: S65599  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-15; 16-33 <SOE>  
C:Superfamily: hevein precursor; barwin homology; hevein chitin-binding domain homology

Query Match 100.0%; Score 19; DB 2; Length 33;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
|||  
Db 20 FNP 22

RESULT 13  
S42876  
probable succinate--CoA lyase (ADP-forming) (EC 6.2.1.5) beta chain - Coxiella burnetii  
N:Alternate names: probable succinyl-CoA synthetase beta chain

C:Species: *Coxiella burnetii*  
C:Date: 20-Feb-1995 #sequence\_revision 26-Jul-1996 #text\_change 05-May-2000  
C:Accession: S42876  
R:Title: D.; Willems, H.; Oswald, W.; Krauss, H.  
submitted to the EMBL Data Library, March 1994  
A:Reference number: S42872  
A:Accession: S42876  
A:Molecule type: DNA  
A:Residues: 1-35 <THI>  
A:Cross-references: EMBL:X77919  
C:Superfamily: succinate--CoA ligase (ADP-forming) beta chain  
C:Keywords: acid-thiol ligase; coenzyme A

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
DB 26 FNP 28

RESULT 14  
H82103  
hypothetical protein VC2228 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)  
C:Species: *Vibrio cholerae*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: H82103  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833  
A:Accession: H82103  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-35 <HEI>  
A:Cross-references: GB:AE04294; GB:AE03852; NID:96656774; PIDN:AAF95372.1; GSPDB:GN001  
C:Genetics:  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Gene: VC2228  
A:Map position: 1

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
DB 4 FNP 6

RESULT 15  
A55850  
7alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) - *Clostridium sordeilii*  
C:Species: *Clostridium sordeilii*  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 15-Oct-1999  
C:Accession: A55850  
R:Coleman, J.P.; Hudson, L.L.; Adams, M.J.  
J. Bacteriol. 176, 4865-4874, 1994  
A:Title: Characterization and regulation of the NADP-linked 7alpha-hydroxysteroid dehydro-  
A:Reference number: A55850; MUID:94327451  
A:Accession: A55850  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-35 <COL>  
A:Cross-references: GB:L12058; NID:9431280; PIDN:AAA5355.1; PID:9577669  
C:Keywords: oxidoreductase

Query Match 100.0%; Score 19; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
DB 12 FNP 14

Search completed: August 20, 2002, 11:10:41  
Job time: 5671 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:33:03 ; Search time 55.29 Seconds

(Without alignments)  
2.101 Million cell updates/sec

Title: US-09-824-286-17

Perfect score: 19

Sequence: 1 FNP 3

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	15	ALIS_MANSE	P42559 manduca sex
2	19	100.0	15	RM12_YEAST	P36522 saccharomyc
3	19	100.0	39	ABAE_BOMPA	P81463 bombus pasc
4	19	100.0	40	PSAJ_CYAPA	P48117 cyanophora
5	19	100.0	46	COAT_BPH75	P82889 bacterioph
6	19	100.0	46	YPCA_ECOLI	P19755 escherichia
7	19	100.0	50	RL39_SUDAC	P13005 sulfolobus
8	19	100.0	51	INS_ANGRO	P42633 anguilla ro
9	19	100.0	52	MTK_DROME	P24395 drosophila
10	19	100.0	53	ABAE_APIME	P15450 apis mellif
11	19	100.0	55	ITRP_HALRO	P16589 halocynthia
12	19	100.0	61	PSBK_HORVU	P25877 hordeum vul
13	19	100.0	61	PSBK_MAIZE	P48188 zea mays (m
14	19	100.0	61	PSBK_ORYSA	P12162 oryza sativ
15	19	100.0	63	PSBK_WHEAT	P58273 triticum ae
16	19	100.0	63	RPCX_SCHPO	P48011 schizosacch
17	19	100.0	64	Y592_HAEIN	P44021 haemophilus
18	19	100.0	66	ATP8_CEREH	Q2mq19 cervus elap
19	19	100.0	68	YE02_HAEIN	P44177 haemophilus
20	19	100.0	70	CSP1_ECOLI	P77605 escherichia
21	19	100.0	70	CSP1_SALTI	P59876 salmonella
22	19	100.0	74	CSPJ_SALTY	P32681 gallus gall
23	19	100.0	74	HXB8_CHICK	O62425 mus musculu
24	19	100.0	75	NDML_MOUSE	P01054 hordeum vul
25	19	100.0	77	ICIC_HORVU	P57474 buchnera ap
26	19	100.0	79	RS16_BUCAI	P34953 macaca fasc
27	19	100.0	81	IAC_MACFA	O00483 homo sapien
28	19	100.0	81	NDML_HUMAN	O01331 bos taurus
29	19	100.0	82	NUML_BOVIN	P02372 escherichia
30	19	100.0	82	RS16_ECOLI	P44382 haemophilus
31	19	100.0	82	RS16_HAEIN	P58123 pasteurella
32	19	100.0	82	RS16_PASMU	P36242 salmonella
33	19	100.0	82	RS16_SALTY	

34	19	100.0	82	RS16_VIBCH	O9kug0 vibrio chol
35	19	100.0	82	SIRA_VIBCH	O9kvw4 vibrio chol
36	19	100.0	83	ICIB_HORVU	P10663 hordeum vul
37	19	100.0	83	RS16_PSRAE	O9hxp9 pseudomonas
38	19	100.0	84	IAC2_HUMAN	P20155 homo sapien
39	19	100.0	86	ILVM_ECOLI	P13048 escherichia
40	19	100.0	86	YIDD_PROMI	P22834 proteus mir
41	19	100.0	88	VG30_BPM2	O64224 mycobacteri
42	19	100.0	88	YVAD_VACCC	P20513 vaccinia vl
43	19	100.0	92	HXB8_RAT	P18863 rattus norv
44	19	100.0	92	Y766_RICPR	O9zch6 ticketia
45	19	100.0	94	F16B_MOUSE	P97323 mus musculu

## ALIGNMENTS

RESULT 1	ALIS_MANSE	STANDARD	PRT	15 AA.
ID	ALIS_MANSE			
AC	P42559			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Allatostatin (Mas-AS).			
OS	Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;			
OC	Sphingioidea; Sphingidae; Sphinginae; Manduca.			
OX	NCBI_TaxID=7130;			
RN	[1]			
RP	SEQUENCE.			
RC	TISSUE=Head;			
RX	MEDLINE=92052112; PubMed=1946359;			
RA	Kramer S.-J., Toschl A., Miller C.A., Kataoka H., Quistad G.B.,			
RA	L.J.P., Carney R.L., Schooley D.A.;			
RT	"Identification of an allatostatin from the tobacco hornworm Manduca			
RT	sexta.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 88:9458-9462(1991).			
CC	-1- FUNCTION: STRONGLY INHIBITS JUVENILE HORMONE BIOSYNTHESIS IN VITRO			
CC	BY THE CORPORA ALLATA FROM FIFTH-STADIUM LARVAE AND ADULT FEMALES.			
CC	-1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.			
KW	Neuropeptide.			
FT	MOD_RES			
FT	MOD_RES 1 1 PYROGLUTAMIC CARBOXYLIC ACID.			
SQ	SEQUENCE 15 AA: 1908 MW: 1605B77CDBBC838E CRC64;			
Query Match 100.0%; Score 19; DB 1; Length 15;				
Best Local Similarity 100.0%; Pred. No. 78;				
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 FNP 3			
Db	9 FNP 11			
RESULT 2				
ID	RM12_YEAST	STANDARD;	PRT;	15 AA.
AC	P36522;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-JUN-1994 (Rel. 29, Last annotation update)			
DE	Mitochondrial 60S ribosomal protein L12 (ImL12) (Fragment).			
GN	MRPL12.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE.			
RX	MEDLINE=91285106; PubMed=2060626;			
RA	Grohmann L., Graack H.-R., Kruff V., Choll T., Goldschmidt-Reisin S.,			

```

RA Kitakawa M.;
RT "extended N-terminal sequencing of proteins of the large ribosomal
RL subunit from yeast mitochondria.";
RL FEBS Lett. 284:51-56(1991).
DR PIR: S17261; S17261.
DR SGD: L0002667; MRPL12.
KM Ribosomal protein; Mitochondrion.
FT NON_TER
SQ SEQUENCE 15 AA; 1851 MW; 74BCD9FEDDB3900 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
DB 13 FNP 15

RESULT 3
ABAE_BOMPA
ID ABAE_BOMPA STANDARD; PRT; 39 AA.
AC P81463;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Abaecin.
OS Bombus pascuorum.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Prexygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Bombs.
OX NCBI_TaxID=65598;
RN [1]
RP SEQUENCE.
RC TISSUE=Hemolymph;
RX MEDLINE=87362903; PubMed=9219367;
RA Rees J.A., Montalvo M., Bulet P.;
RT "Novel antibacterial peptides isolated from a European bumblebee,
RL Bombus pascuorum (Hymenoptera, Apoidea).";
RL Insect Biochem. Mol. Biol. 27:413-422(1997).
CC -1- FUNCTION: ANTIBACTERIAL PEPTIDE AGAINST GRAM-POSITIVE AND
CC GRAM-NEGATIVE BACTERIA.
CC -1- INDUCTION: BY BACTERIAL INFECTION.
CC -1- SIMILARITY: PARTIAL TO APIDAEINS AND DIPTERICINS.
KM Insect immunity; Antibiotic; Hemolymph.
SQ SEQUENCE 39 AA; 4395 MW; 52E952E25D13A097 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
DB 24 FNP 26

RESULT 4
PSAJ_CYAPA
ID PSAJ_CYAPA STANDARD; PRT; 40 AA.
AC P84117;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Photosystem I reaction center subunit IX (PSI-J).
GN PSAJ.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=LB555 / PRINGSHEIM;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / PRINGSHEIM;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -1- FUNCTION: MAY HELP IN THE ORGANIZATION OF THE PSAE AND PSAF
CC SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE PSAJ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U00821; AAA81183.1; -
DR Mendel: 7884; CYAPA:psaj.1.
DR InterPro: IPR002615; PSI_PSAJ.
DR Pfam: PF01701; PSI_PSAJ.1.
DR ProDom: PD004198; PSI_PSAJ.1.
KW Cyanelle; Photosystem I; Photosynthesis; Transmembrane.
FT TRANSMEM
SQ SEQUENCE 40 AA; 4481 MW; 798A8CA36A007153 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3
DB 31 FNP 33

RESULT 5
COAT_BPH75
ID COAT_BPH75 STANDARD; PRT; 46 AA.
AC P82889;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein P8 (Major coat protein).
OS Bacteriophage PH75.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxID=144736;
RN [1]
RP SEQUENCE, AND FORMYLATION.
RX MEDLINE=21265138; PubMed=11371161;
RA Pederson D.M., Welsh L.C., Marvin D.A., Sampson M., Perham R.N.,
RA Yu M., Slater M.R.;
RT "The protein capsid of filamentous bacteriophage PH75 from Thermus
RT thermophilus.";
RL J. Mol. Biol. 309:401-421(2001).
CC -1- FUNCTION: COAT PROTEIN P8 IS THE MAJOR COAT PROTEIN OF THE
CC VIRION.
CC -1- SUBUNIT: THERE ARE SEVERAL COPIES OF THIS PROTEIN IN THE COAT OF
CC PHAGE PH75. THEY ARE ARRANGED IN A HELIX AROUND THE DNA.
CC -1- SUBCELLULAR LOCATION: Type III membrane protein (Potential).
KW Coat protein; Formylation; Transmembrane.

```



FT DOMAIN 1 15 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT DOMAIN 37 46 INTRACELLULAR (POTENTIAL).  
 FT MOD\_RES 1 1 FORMYLATION.  
 SQ SEQUENCE 46 AA; 4813 MW; 1E2BFA950C7ACBE CRC64;

Query Match 100.0%; Score 19; DB 1; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
 |||  
 Db 3 FNP 5

RESULT 6  
 YPC4\_ECOLI STANDARD; PRT; 46 AA.  
 ID YPC4\_ECOLI  
 AC P19755;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-NOV-1991 (Rel. 20, Last annotation update)  
 DE Hypoetical 5.3 kDa protein (ORF 46).  
 OS Escherichia coli.  
 CC Escherichia coli.  
 CC Plasmid Incn PCU1.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia  
 CC NCBI\_Taxid=562;  
 RN 11)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90382682; PubMed=2205534;  
 RA Krishnan B.R., Robert P.R., Seltzer U., Iyer V.N.;  
 RT "Mutations within the replicon of the Incn plasmid PCU1 that affect  
 its Escherichia coli polA-independence but not its autonomous  
 RT replication ability.";  
 RL Gene 91:1-7(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M18262; AAA98067.1; -;  
 DR PIR; S12084; S12084.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 46 AA; 5304 MW; 14F81E5F636952B2 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 46;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
 |||  
 Db 35 FNP 37

RESULT 7  
 RL39\_SULAC STANDARD; PRT; 50 AA.  
 ID RL39\_SULAC  
 AC P13005; P95686;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 50S ribosomal protein L39e (L46e).  
 GN RPL39E OR RPL46 OR RIL46.  
 OS Sulfolobus acidocaldarius.  
 CC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.

OX NCBI\_Taxid=2285;

RN 11)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-46.  
 RC STRAIN=ATCC 35091 / DSM 1616 / PI;  
 RX MEDLINE=89325607; PubMed=2502431;  
 RA Ramirez C., Louie K.A., Matheson A.T.;  
 RT "A small basic ribosomal protein in Sulfolobus solfataricus  
 RT equivalent to L46 in yeast: structure of the protein and its gene.";  
 RL FEBS Lett. 250:416-418(1989).

RN 12)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;  
 RX MEDLINE=95226466; PubMed=7711082;  
 RA Moll R., Schmidtke S., Schaefer G.;  
 RT "Nucleotide sequence of a gene cluster encoding ribosomal proteins in  
 RT the thermophilic crenarchaeon Sulfolobus acidocaldarius.";  
 RL Biochim. Biophys. Acta 1261:315-318(1995).  
 CC -1- FUNCTION: BINDS SPECIFICALLY TO A REGION IN 26S RNA.  
 CC NEAR THE SUBUNIT INTERFACE.  
 CC -1- SIMILARITY: BELONGS TO THE L39E FAMILY OF RIBOSOMAL PROTEINS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X16161; CAA34287.1; -;  
 DR EMBL; X77509; CAA54638.1; ALT\_INIT.  
 DR PIR; S05009; R6UC46.  
 DR PIR; S41963; S41963.  
 DR InterPro; IPR000077; Ribosomal\_L39.  
 DR Pfam; PF00832; Ribosomal\_L39; 1.  
 DR Prodom; PD007914; Ribosomal\_L39; FALSE\_NEG.  
 DR PROSITE; PS00051; RIBOSOMAL\_L39E; 1.  
 KW Ribosomal protein.  
 FT INT MET 0  
 SQ SEQUENCE 50 AA; 5932 MW; 217CB51B357B44F9 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
 |||  
 Db 36 FNP 38

RESULT 8  
 INS\_ANGRO STANDARD; PRT; 51 AA.  
 ID INS\_ANGRO  
 AC P42633;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insulin.  
 GN INS.  
 OS Anguilla rostrata (American eel).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;  
 CC Anguillidae; Anguilla.  
 CC NCBI\_Taxid=7938;  
 RN 11)  
 RP SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=91340068; PubMed=1874385;  
 RA Conlon J.M., Andrews P.C., Thim L., Moon T.W.;  
 RT "The primary structure of glucagon-like peptide but not insulin has  
 RT been conserved between the American eel, Anguilla rostrata and the  
 RT European eel, Anguilla anguilla.";

FL Gen. Comp. Endocrinol. 82:23-32(1991).  
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT  
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND  
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE  
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.  
CC -1- SUBUNIT: HEMODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO  
CC DISULFIDE BONDS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.  
DR HSSP: P01308; 1HIS.  
DR InterPro: IPR000739; Insulin\_IGF\_Relaxin.  
DR Pfam: PF000049; Insulin\_1.  
DR PRINTS: PR00276; INSULIN.  
DR PRINTS: PR00277; INSULIN.  
DR SMART: SM00078; 1IGF; 1.  
DR PROSITE: PS00262; INSULIN; 1.  
KV Insulin family; Hormone; Glucose metabolism.  
FT CHAIN 1 30 INSULIN B CHAIN.  
FT NON CONS 30 31  
FT CHAIN 31 51 INSULIN A CHAIN.  
FT DISULFID 7 37  
FT DISULFID 19 50 INTERCHAIN.  
FT DISULFID 36 41  
SQ SEQUENCE 51 AA; 5652 MW; 1999FD7EBA173CB2 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
Db 26 FNP 28

RESULT 9  
MTK\_DROME STANDARD; PRT: 52 AA.  
AC Q24395; Q24396; Q9Y7B9;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Metchnikowin precursor.  
GN MTK OR CG8175.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 27-52.  
RC STRAIN=OREGON-R; TISSUE=Abdomen, and Thorax;  
RX MEDLINE=96067716; PubMed=758819;  
RA Levashina E.A., Ohresser S., Bulet B., Reichhart J.-M., Hetru C.,  
RA Hoffmann J.A.;  
RT "Metchnikowin, a novel immune-inducible proline-rich peptide from  
RT Drosophila with antibacterial and antifungal properties.";  
RL Eur. J. Biochem. 233:694-700(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-R;  
RX MEDLINE=98263241; PubMed=9600835;  
RA Levashina E.A., Ohresser S., Lemaitre B., Imler J.-L.;  
RT "Two distinct pathways can control expression of the gene encoding  
RT the Drosophila antimicrobial peptide metchnikowin.";  
RL J. Mol. Biol. 278:515-527(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck H., Brockstein P., Brottier A.,  
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies S.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: POTENT ANTIFUNGAL AND ANTIBACTERIAL ACTIVITY AGAINST  
CC GRAM-POSITIVE BACTERIA.  
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN FAT BODY. MAYBE BLOOD  
CC CELLS PARTICIPATE IN THE PRODUCTION OF THE PEPTIDE.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED RAPIDLY AND STRONGLY AT ALL STAGES.  
CC -1- POLYMORPHISM: 2 ALLELIC FORMS (A1 AND A2) VARYING IN TWO AA  
CC POSITION. THE ISOFORM SHOWN HERE IS A1.  
CC -1- SIMILARITY: TO DIPERICIN, HEMIPERICIN, DNOSOCIN, APIDECINS AND  
CC TO THE C-TERMINUS OF ABAECIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X91060; CAA62511.1; -;  
DR EMBL: X91061; CAA62512.1; -;  
DR EMBL: AF030959; AAC64659.1; -;  
DR EMBL: AE003811; AAF58139.1; -;  
DR FLYbase: FBgn0014865; Mtk.  
KW Insect immunity; Antibiotic; Fungicide; Signal.  
FT SIGNAL 1 24  
FT PROPEP 25 26  
FT PEPTIDE 27 52  
FT VARIANT 19 19  
FT VARIANT 29 29  
SQ SEQUENCE 52 AA; 5654 MW; 0B1E2112BA803129 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 52;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FNP 3  
Db 26 FNP 28

Db 41 FNP 43

RESULT 10  
ID ABAE\_APIME STANDARD; PRT; 53 AA.  
AC P15450;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Abaecin precursor.  
OS Apis mellifera (Honeybee).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
OC Apoidea; Apidae; Apis.  
NCBI\_TaxID=7460;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95050655; Pubmed=7961803;  
RA Casteels-Josson K., Zhang W., Capaci T., Tempst P.;  
RT "Acute transcriptional response of the honeybee peptide-antibiotics  
RT gene repertoire and required post-translational conversion of the  
RT precursor structures.";  
RT J. Biol. Chem. 269:28569-28575(1994).  
RN [2]  
RP SEQUENCE OF 20-53.  
RC TISSUE=Hemolymph;  
RX MEDLINE=90126848; Pubmed=2298215;  
RA Casteels P., Ampe C., Riviere L., van Damme J., Elicone C., Jacobs F.,  
RA Tempst P.;  
RT "Isolation and characterization of abaecin, a major antibacterial  
RT response peptide in the honeybee (Apis mellifera).";  
RL Eur. J. Biochem. 187:381-386(1990).  
CC -1- FUNCTION: THIS PEPTIDE HAS BACTERICIDAL ACTIVITY.  
CC -1- SIMILARITY: PARTIAL TO APIDACINS AND DIPERICINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U15954; AAA67442.1; -  
DR PIR: S08152; S08152.  
KW Insect immunity; Antibiotic; Hemolymph; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 53 ABAECIN.  
FT VARIANT 52 52 G->S.  
SQ SEQUENCE 53 AA; 5903 MW; 0EB1F0CC57797FBC CRC64;

Query Match 100.0%; Score 19; DB 1; Length 53;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3  
|||  
Db 43 FNP 45

RESULT 11  
ID ITRP\_HALRO STANDARD; PRT; 55 AA.  
AC P16589;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Trypsin inhibitor.  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyridae; Halocynthia.

OX NCBI\_TaxID=7729;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Hemolymph;  
RX MEDLINE=90256715; Pubmed=2341375;  
RA Kunazaki T., Hoshida N., Yokosawa H., Ishii S.-I.;  
RT "Primary structure of ascidian trypsin inhibitors in the hemolymph of  
RT a solitary ascidian, Halocynthia roretzi.";  
RL J. Biochem. 107:409-413(1990).  
RN [2]  
RP DISULFIDE BONDS.  
RX MEDLINE=90256716; Pubmed=2111316;  
RA Kunazaki T., Ishii S.-I.;  
RT "Disulfide bridge structure of ascidian trypsin inhibitor I:  
RT similarity to Kazal-type inhibitors.";  
RL J. Biochem. 107:414-419(1990).  
CC -1- FUNCTION: POTENT INHIBITOR OF TRYPSIN.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: SOME TOPOLOGICAL SIMILARITY TO KAZAL-TYPE INHIBITORS.  
KW Serine protease inhibitor; Hemolymph.  
FT DISULFID 5 40  
FT DISULFID 12 41  
FT DISULFID 14 37  
FT DISULFID 23 34  
FT ACT\_SITE 16 17 REACTIVE BOND.  
SQ SEQUENCE 55 AA; 6078 MW; 412154BDB33207C2 CRC64;

Query Match 100.0%; Score 19; DB 1; Length 55;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FNP 3  
|||  
Db 8 FNP 10

RESULT 12  
ID PSBK\_HORVU STANDARD; PRT; 61 AA.  
AC P25877;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Photosystem II reaction center protein K precursor (PSII-K).  
GN PSBK.  
OS Hordeum vulgare (Barley), and  
OS Secale cereale (Rye).  
OS Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
OC Triticeae; Hordeum.  
NCBI\_TaxID=4513, 4550;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=H. vulgare; TISSUE=Seedling;  
RX MEDLINE=90291518; Pubmed=1694111;  
RA Sexton T.B., Jones J.T., Mulliet J.E.;  
RT "Sequence and transcriptional analysis of the barley ctDNA region  
RT upstream of psbD-psbC encoding trnK(UUU), trpS16, trnD(UUC), psbK,  
RT psbI, and trnS(GCU).";  
RL Curr. Genet. 17:445-454(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S. cereale; Pubmed=2644624;  
RX MEDLINE=89128450;  
RA Burkhart A.A., Kolosov V.L., Klesovich O.N., Zolotarev A.S.;  
RT "Nucleotide sequence of rye chloroplast DNA fragment, comprising  
RT psbD, psbC and trnS genes.";  
RL Nucleic Acids Res. 17:798-798(1989).  
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER  
CC OF PHOTOSYSTEM II.  
CC -1- SIMILARITY: BELONGS TO THE PSBK FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X52765; CAA36975.1; -  
DR EMBL: X61674; CAA43848.1; -  
DR PIR: S28768; S28768.  
DR PIR: S31447; S31447.  
DR InterPro: IPR003687; PsbK.  
DR Pfam: PF02533; PsbK; 1.  
KW Photosystem II; Chloroplast.  
FT PROPEP 1 24  
FT CHAIN 25 61  
SQ SEQUENCE 61 AA; 6996 MW; 84C901067C454777 CRC64;  
-----  
Query Match 100.0%; Score 19; DB 1; Length 61;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 1 FNP 3  
DB 33 FNP 35  
-----  
RESULT 13  
PSBK\_MAIZE STANDARD; PRT; 61 AA.  
ID PSBK\_MAIZE  
AC P48188;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Photosystem II reaction center protein K precursor (PSII-K).  
GN PSBK.  
OS Zea mays (Maize).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoidae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95395841; PubMed=766415;  
RA Maier R.M., Neckermann K., Igloi G.L., Koessel H.;  
RT "Complete sequence of the maize chloroplast genome: gene content,  
RT hotspots of divergence and fine tuning of genetic information by  
RT transcript editing";  
RL J. Mol. Biol. 251:614-628(1995).  
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER  
CC OF PHOTOSYSTEM II.  
CC -1- SIMILARITY: BELONGS TO THE PSBK FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X86563; CAA60268.1; -  
DR MaizeDB: 118217; -  
DR Mendel: 2686; ZEma;psbk;1.  
DR InterPro: IPR003687; PsbK.  
DR Pfam: PF02533; PsbK; 1.  
KW Photosystem II; Chloroplast.  
FT PROPEP 1 24  
FT CHAIN 25 61  
SQ SEQUENCE 61 AA; 6982 MW; BHC361067C4158P5 CRC64;  
-----  
Query Match 100.0%; Score 19; DB 1; Length 61;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 1 FNP 3  
DB 33 FNP 35  
-----  
RESULT 15  
PSBK\_WHEAT STANDARD; PRT; 61 AA.  
ID PSBK\_WHEAT

-----  
SQ SEQUENCE 61 AA; 6874 MW; D67361066179A2C1 CRC64;  
-----  
Query Match 100.0%; Score 19; DB 1; Length 61;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 1 FNP 3  
DB 33 FNP 35  
-----  
RESULT 14  
PSBK\_ORYZA STANDARD; PRT; 61 AA.  
ID PSBK\_ORYZA  
AC P12162;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Photosystem II reaction center protein K precursor (PSII-K).  
GN PSBK.  
OS Oryza sativa (Rice).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RX MEDLINE=89364698; PubMed=2770692;  
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,  
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,  
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,  
RA Kanno A., Nishizawa Y., Hirai A., Shinozaki K., Sugita M.;  
RT "The complete sequence of the rice (Oryza sativa) chloroplast genome:  
RT intermolecular recombination between distinct trna genes accounts for  
RT a major plastid DNA inversion during the evolution of the cereals";  
RL Mol. Gen. Genet. 217:185-194(1989).  
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER  
CC OF PHOTOSYSTEM II.  
CC -1- SIMILARITY: BELONGS TO THE PSBK FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X15901; CAA34010.1; -  
DR PIR: J00203; F2R2KS.  
DR Mendel: 2688; ORYsa;psbk;1.  
DR InterPro: IPR003687; PsbK.  
DR Pfam: PF02533; PsbK; 1.  
KW Photosystem II; Chloroplast.  
FT PROPEP 1 24  
FT CHAIN 25 61  
SQ SEQUENCE 61 AA; 6982 MW; BHC361067C4158P5 CRC64;  
-----  
Query Match 100.0%; Score 19; DB 1; Length 61;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
QY 1 FNP 3  
DB 33 FNP 35  
-----  
RESULT 15  
PSBK\_WHEAT STANDARD; PRT; 61 AA.  
ID PSBK\_WHEAT

AC P58273;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Photosystem II reaction center protein K precursor (PSII-K).  
 GN PSBK.  
 OS *Triticum aestivum* (Wheat).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; *Triticum*.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPRAIN-cv. Chinese Spring;  
 RA Ogihara Y., Isono K., Kojima T., Endo A., Hanaoka M., Shina T.,  
 RA Terachi T., Utsugi S., Murata M., Mori N., Takumi S., Ikeo K.,  
 RA Gojobori T., Murai R., Murai K., Matsuoka Y., Ohnishi Y., Tajiri H.,  
 RA Tsunewaki K.;  
 RT "Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome:  
 RT complete sequence and contig clones.";  
 RL Plant Mol. Biol. Rep. 18:243-253(2000)  
 CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER  
 CC OF PHOTOSYSTEM II.  
 CC -!- SIMILARITY: BELONGS TO THE PSBK FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB042240; BAF47017.1; -;  
 DR InterPro: IPR003687; PSBK.  
 DR Pfam: PF02533; PSBK; 1.  
 KW Photosystem II; Chloroplast.  
 FT PROPEP 1 24 BY SIMILARITY.  
 FT CHAIN 25 61 PHOTOSYSTEM II REACTION CENTER PROTEIN K.  
 SO SEQUENCE 61 AA; 7030 MW; 467901067C4158FD CRC64;

Query Match 100.0%; Score 19; DB 1; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
 |||  
 DB 33 FNP 35

Search completed: August 20, 2002, 11:33:04  
 Job time: 1453 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 11:31:58 ; Search time 191.14 Seconds

(Without alignments)  
2.715 Million cell updates/sec

Title: US-09-824-286-17

Perfect score: 19  
Sequence: 1 FNP 3

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertedrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvrius:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	11	12	084073
2	19	100.0	16	6	Q9TOY6
3	19	100.0	18	5	Q9TWM9
4	19	100.0	18	10	Q9S817
5	19	100.0	20	2	Q9R509
6	19	100.0	20	5	Q25281
7	19	100.0	20	6	P82666
8	19	100.0	21	5	Q9TWM8
9	19	100.0	22	4	Q96024
10	19	100.0	22	12	Q64831
11	19	100.0	23	10	Q9S821
12	19	100.0	25	2	Q32326
13	19	100.0	25	11	Q63988
14	19	100.0	27	5	P90722
15	19	100.0	27	6	Q9XSC0
16	19	100.0	27	12	Q68908

17	19	100.0	27	12	Q91L61	Q91161 hepatitis c
18	19	100.0	27	12	Q91L60	Q91160 hepatitis c
19	19	100.0	27	12	Q91L59	Q91159 hepatitis c
20	19	100.0	27	12	Q91L58	Q91158 hepatitis c
21	19	100.0	27	12	Q91L57	Q91157 hepatitis c
22	19	100.0	27	12	Q91L56	Q91156 hepatitis c
23	19	100.0	27	12	Q91L55	Q91155 hepatitis c
24	19	100.0	27	12	Q91L54	Q91154 hepatitis c
25	19	100.0	27	12	Q91L53	Q91153 hepatitis c
26	19	100.0	27	12	Q91L52	Q91152 hepatitis c
27	19	100.0	27	12	Q91L51	Q91151 hepatitis c
28	19	100.0	27	12	Q91L50	Q91150 hepatitis c
29	19	100.0	27	12	Q91L49	Q91149 hepatitis c
30	19	100.0	27	12	Q91L48	Q91148 hepatitis c
31	19	100.0	27	12	Q91L47	Q91147 hepatitis c
32	19	100.0	27	12	Q91L46	Q91146 hepatitis c
33	19	100.0	27	12	Q91L45	Q91145 hepatitis c
34	19	100.0	27	12	Q91L44	Q91144 hepatitis c
35	19	100.0	27	12	Q91L43	Q91143 hepatitis c
36	19	100.0	27	12	Q91L42	Q91142 hepatitis c
37	19	100.0	27	12	Q91L41	Q91141 hepatitis c
38	19	100.0	27	12	Q91L40	Q91140 hepatitis c
39	19	100.0	27	12	Q91L39	Q91139 hepatitis c
40	19	100.0	27	12	Q91L38	Q91138 hepatitis c
41	19	100.0	27	12	Q91L37	Q91137 hepatitis c
42	19	100.0	27	12	Q91L36	Q91136 hepatitis c
43	19	100.0	27	12	Q91L35	Q91135 hepatitis c
44	19	100.0	27	12	Q91L34	Q91134 hepatitis c
45	19	100.0	27	12	Q91L33	Q91133 hepatitis c

#### ALIGNMENTS

RESULT 1  
ID Q84073 PRELIMINARY; PRT; 11 AA.  
AC Q84073;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE INFLUENZA A/EPV/ROSTOCK/34 (H7N1), POLYMERASE 3 (SEG 3), 3' END OF  
DE VNNA (INITIATOR REGION FOR PROTEIN CODING) (FRAGMENT).  
OS Influenza A virus.  
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza virus A and B group; Influenza A viruses.  
OX NCBI\_TaxID=11320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=80034428; PubMed=493121;  
RX Robertson J.S.;  
RT "5' and 3' terminal nucleotide sequences of the rna genome segments of  
RL Influenza virus.",  
RL Nucleic Acids Res. 6:3745-3757(1979).  
DR EMBL: J02123; AAA43612.1; -.  
FT NON\_TER 11 11  
SQ SEQUENCE 11 AA; 1400 MW; CC2007F7AC412C9 CRC64;

Query Match 100.0%; Score 19; DB 12; Length 11;  
Best local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNP 3  
Db 9 FNP 11  
RESULT 2  
ID Q9TOY6 PRELIMINARY; PRT; 16 AA.  
AC Q9TOY6;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE ANNEXIN I (FRAGMENT).  
 OC Oryctolagus cuniculus (Rabbit).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=96226532; PubMed=8967522;  
 RA Mayran N., Traverso V., Maroux S., Massey-Harroche D.;  
 RT "Cellular and subcellular localizations of annexin I, IV, and VI in  
 lung epithelia.";  
 RL Am. J. Physiol. 270:L863-L871(1996).  
 SO SEQUENCE 16 AA; 1546 MW; 70AA0A45193C8BB6 CRC64;

Query Match 100.0%; Score 19; DB 6; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 DB 14 FNP 16

RESULT 3  
 O9TWM9 PRELIMINARY; PRT; 18 AA.  
 AC O9TWM9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE NTAG ANTIGEN (FRAGMENT).  
 CC Tetrahymena pyriformis.  
 CC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 CC Tetrahymenina; Tetrahymena.  
 OX NCBI\_TaxID=5908;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=94323706; PubMed=7519359;  
 RA Leary J.H.3rd, Evans D.L., Jaso-Friedmann L.;  
 RT "Partial amino acid sequence of a novel protozoan parasite antigen  
 that inhibits non-specific cytotoxic cell activity.";  
 RL Scand. J. Immunol. 40:158-164(1994).  
 SO SEQUENCE 18 AA; 2082 MW; 13BCB737B1FF92A3 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 DB 7 FNP 9

RESULT 4  
 O9S817 PRELIMINARY; PRT; 18 AA.  
 AC O9S817;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE 40 KDA pI 8.5 ABSCTISSIC ACID-INDUCED HISTIDINE RICH PROTEIN  
 (FRAGMENT).  
 CC Oryza sativa (Rice).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE.

RX MEDLINE=95175599; PubMed=7870812;  
 RA Moons A., Bauw G., Prinsen E., Van Montagu M., Van der Straeten D.;  
 RT "Molecular and physiological responses to abscisic acid and salts in  
 roots of salt-sensitive and salt-tolerant indica rice varieties.";  
 RL Plant Physiol. 107:177-186(1995).  
 SO SEQUENCE 18 AA; 2094 MW; 0CD245DB237E7520 CRC64;

Query Match 100.0%; Score 19; DB 10; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 DB 4 FNP 6

RESULT 5  
 O9R509 PRELIMINARY; PRT; 20 AA.  
 AC O9R509;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ADENYLATE KINASE (EC 2.7.4.3) (FRAGMENT).  
 CC Bacillus alcalophilus.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1445;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=94117390; PubMed=8288548;  
 RA Gilles A.M., Glaser P., Perrier V., Meier A., Longin R., Sebald M.,  
 RA Maignan L., Pistotnik E., Barzu O.;  
 RT "Zinc, a structural component of adenylate kinases from gram-positive  
 bacteria.";  
 RL J. Bacteriol. 176:520-523(1994).  
 DR InterPro:IPR000850; Adenylate\_kin.  
 DR Pfam:PF00406; adenylatekinase; 1.  
 SO SEQUENCE 20 AA; 2190 MW; 94D98FC289220254 CRC64;

Query Match 100.0%; Score 19; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 DB 6 FNP 8

RESULT 6  
 Q25281 PRELIMINARY; PRT; 20 AA.  
 AC Q25281;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALPHA-TUBULIN DNA (FRAGMENT).  
 CC Leishmania enriettii.  
 CC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5663;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84142075; PubMed=6321982;  
 RA Wirth D.F., Slater C.;  
 RT "Isolation and characterization of an alpha-tubulin gene from  
 Leishmania enriettii.";  
 RL Mol. Biochem. Parasitol. 9:83-92(1983).  
 DR EMBL: M28001; AAA29273.1;  
 DR InterPro:IPR003008; Tubulin\_FtsZ.  
 DR Pfam:PF00091; tubulin; 1.  
 KW GTP-binding.



FT NON\_TER 1 1  
 FT NON\_TER 20 20  
 SQ SEQUENCE 20 AA; 2351 MW; B9A8F2D1755D810D CRC64;

Query Match 100.0%; Score 19; DB 5; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 |||  
 Db 17 FNP 19

RESULT 7  
 P82666 PRELIMINARY; PRT; 20 AA.  
 ID P82666;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)  
 DE MITOCHONDRIAL 28S RIBOSOMAL PROTEIN S23 (MRP-S23) (FRAGMENTS).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 OX [1]  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=LIVER;  
 RX MEDLINE=20490686; PubMed=10938081;  
 RA Koc E.C., Burkhardt W., Blackburn K., Moseley A., Koc H.,  
 RA Spremlj L.L.;  
 RT "A proteomics approach to the identification of mammalian  
 RT mitochondrial small subunit ribosomal proteins.";  
 RL J. Biol. Chem. 275:32585-32591(2000).  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.  
 KW Ribosomal protein; Mitochondrion.  
 FT NON\_TER 1 1  
 FT NON\_CONS 10 11  
 FT NON\_TER 20 20  
 SQ SEQUENCE 20 AA; 2249 MW; EC7FE3CA50071EE4 CRC64;

Query Match 100.0%; Score 19; DB 6; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 |||  
 Db 15 FNP 17

RESULT 8  
 O9TWM8 PRELIMINARY; PRT; 21 AA.  
 ID O9TWM8;  
 AC O9TWM8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)  
 DE NKTAG ANTIGEN (FRAGMENT).  
 OS Tetrahymena pyriformis.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 OC Tetrahymenida; Tetrahymena.  
 OC NCBI\_TaxID=5908;  
 OX [1]  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=94323706; PubMed=7519359;  
 RA Leary J.H.3rd, Evans D.L., Jaso-Friedmann L.;  
 RT "Partial amino acid sequence of a novel protozoan parasite antigen  
 RT that inhibits non-specific cytotoxic cell activity.";  
 RL Scand. J. Immunol. 40:158-164(1994).  
 SQ SEQUENCE 21 AA; 2353 MW; E0B7ED0F061127B1 CRC64;

Query Match 100.0%; Score 19; DB 5; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 |||  
 Db 7 FNP 9

RESULT 9  
 O96024 PRELIMINARY; PRT; 22 AA.  
 ID O96024;  
 AC O96024;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE MITOCHONDRIAL RIBOSOMAL PROTEIN S23 (FRAGMENT).  
 GN MRPS23.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429115; PubMed=11543634;  
 RA Kenmochi N., Suzuki T., Uechi T., Magoori M., Higa S.,  
 RA Watanabe K., Tanaka T.;  
 RT "The human mitochondrial ribosomal protein genes: Mapping of 54 genes  
 RT to the chromosomes and implications for human disorders.";  
 RL Genomics 77:65-70(2001).  
 DR EMBL: AB061206; BAB5456.1; -  
 KW Ribosomal protein.  
 FT NON\_TER 1 1  
 FT NON\_TER 22 22  
 SQ SEQUENCE 22 AA; 2536 MW; A079BB04B5E5E7A6 CRC64;

Query Match 100.0%; Score 19; DB 4; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 |||  
 Db 12 FNP 14

RESULT 10  
 O64831 PRELIMINARY; PRT; 22 AA.  
 ID O64831;  
 AC O64831;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, last annotation update)  
 DE I5 FIBER (FRAGMENT).  
 OS Human adenovirus type 5.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OC NCBI\_TaxID=26285;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85092388; PubMed=2981456;  
 RA Cladarras C., Mold W.S.;  
 RT "DNA sequence of the early E3 transcription unit of adenovirus 5.";  
 RL Virology 140:28-43(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85092389; PubMed=3966299;  
 RA Cladarras C., Bhat B., Mold W.S.;  
 RT "Mapping the 5' ends, 3' ends, and splice sites of mRNAs from the  
 RT early E3 transcription unit of adenovirus 5.";  
 RL Virology 140:44-54(1985).  
 DR EMBL: X03002; CAA26788.1; -

FT NON\_TER 22 22  
SQ SEQUENCE 22 AA: 2575 MW: CA394D2EE31FE71C CRC64;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 12; Length 22;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
Db 11 FNP 13

RESULT 11  
OY5821 PRELIMINARY; PRT: 23 AA.  
AC OY5821;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DS LIGHT-HARVESTING ANTENNA OF PHOTOSYSTEM I 680A, LHCI 680A.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=92267013; PubMed=1587270;  
RA Kneetzel J., Svendsen I., Simpson D.J.;  
RT "Identification of the photosystem I antenna polypeptides in barley.  
RT Isolation of three pigment-binding antenna complexes.";  
RL Eur. J. Biochem. 206:209-215(1992).  
SQ SEQUENCE 23 AA: 2282 MW: 194B21969A17581E CRC64;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 10; Length 23;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
Db 17 FNP 19

RESULT 12  
OY32326 PRELIMINARY; PRT: 25 AA.  
AC OY32326;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE REGION UPSTREAM OF THE KDPDE OPERON, KDP A, KDP B AND KDP C GENES.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-DSM 792;  
RX MEDLINE=97369818; PubMed=9226259;  
RA Treuner-Lange A., Kuhn A., Durre P.;  
RT "The kdp system of Clostridium acetobutylicum: cloning, sequencing,  
RT and transcriptional regulation in response to potassium  
RT concentration.";  
RL J. Bacteriol. 179:4501-4512(1997).  
DR EMBL: U44892; AAC45476.1; -  
SQ SEQUENCE 25 AA: 3023 MW: 8D64226018BA2254 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 2; Length 25;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
Db 20 FNP 22

RESULT 13  
OY63988 PRELIMINARY; PRT: 25 AA.  
AC OY63988;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HOX2.4 HOMEOBOX HOMOLOG PROTEIN (FRAGMENT).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE-BONE MORPHOGENETIC PROTEIN-IMPLANTED SUBCUTANEOUS MUSCLE;  
RX MEDLINE=94271262; PubMed=7911662;  
RA Iimura T., Oida S., Takeda K., Maruoka Y., Sasaki S.;  
RT "Changes in homeobox-containing gene expression during ectopic bone  
RT formation induced by bone morphogenetic protein.";  
RT Biochem. Biophys. Res. Commun. 201:980-987(1994).  
DR EMBL: S71284; AAB31004.2; -  
DR InterPro: IPR001356; Homeobox.  
DR Pfam: PF00046; homeobox; 1.  
KW Homeobox; DNA-binding; Nuclear protein.  
FT NON\_TER 1 1  
FT NON\_TER 25 25  
SQ SEQUENCE 25 AA: 3111 MW: 5B88C1D41ACB974 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 11; Length 25;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
Db 2 FNP 4

RESULT 14  
OY90722 PRELIMINARY; PRT: 27 AA.  
AC OY90722;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE HOMEOBOX PROTEIN BHOX 43 (FRAGMENT).  
OS Beroe ovata.  
OC Eukaryota; Metazoa; Ctenophora; Cycliocoela; Beroidea; Beroe.  
OX NCBI\_TaxID=10201;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Witchel H.J., Meech R.W.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U83686; AAB41253.1; -  
DR InterPro: IPR001356; Homeobox.  
DR Pfam: PF00046; homeobox; 1.  
DR PRINTS: PR00024; HOMEOBOX.  
FT NON\_TER 1 1  
FT NON\_TER 27 27  
SQ SEQUENCE 27 AA: 3269 MW: 9861CB88C70F14AC CRC64;

Query Match  
Best Local Similarity 100.0%; Score 19; DB 5; Length 27;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FNP 3  
 Db 2 FNP 4

## RESULT 15

O9XSCO PRELIMINARY; PRT; 27 AA.  
 ID O9XSCO  
 AC O9XSCO;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
 DE BETA-LACTOGLOBULIN C (FRAGMENT).  
 GN LGB.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BREED MERINO-LAND;  
 RA Prinzenberg E.-M., Erhardt G.;  
 RT "Molecular genetic characterization of ovine beta-lactoglobulin C  
 precursor allele and detection by PCR-RFLP.";  
 RL J. Anim. Breed. Genet. 0:0-0(1999).  
 DR EMBL; AF117693; AAD30423.1; -.  
 DR HSSP; P02754; 1CJ5.  
 FT NON\_TER 1  
 SQ SEQUENCE 27 AA; 3051 MW; A4AC05426D1DA34 CRC64;

Query Match 100.0%; Score 19; DB 6; Length 27;  
 Best Local Similarity 100.0%; Pred. NO. 5.8e+02;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 FNP 3  
 Db 16 FNP 18

Search completed: August 20, 2002, 11:31:59  
 Job time: 1473 sec

**THIS PAGE BLANK (USPTO)**